1

SETS

1.1 OPERATIONS ON SETS

The concept of a set is one of the fundamental concepts in mathematics. Set theory permeates most branches of mathematics and yet, in some way, set theory is elusive. For example, if we were to ask for the definition of a set, we may be inclined to give a response such as "it is a collection of objects" or "it is a family of things" and yet the words "collection" and "family" convey no more meaning than the word "set." The reader may be familiar with such a situation in geometry. When we talk of concepts such as points, lines, planes, and distance, we have a general idea of what we are talking about. However at some point in geometry it is necessary to have a list of axioms (the rules that we use in geometry) and definitions (of the main geometrical objects), to deduce theorems about geometry. Nevertheless, some terms must be undefined, although well-understood.

Historically, geometry was the first, best developed, theory based on a system of axioms. However, in secondary school geometry we often study geometric objects without a serious appreciation of the underlying axioms. In a similar way, set theory can also be approached somewhat informally without the kind of rigor that can be established axiomatically. In this book, this approach of using so-called "naive set theory," setting aside sophisticated

An Introduction to Essential Algebraic Structures, First Edition. Martyn R. Dixon, Leonid A. Kurdachenko and Igor Ya. Subbotin.

axiomatic constructions, is the approach we shall use. For us a set will be a collection, class, or system of well-defined and distinct objects of any nature. These objects (the *elements of the set*) are distinct, but altogether they form a new unity, a new whole—a set. We will assume that a set is defined if a rule is given or established, which allows us to determine if an object belongs to the set.

For example, we can define the set of students in the room, the set of computers connected to the Internet in the room now, the set of triangles having a right angle, the set of cars in the parking lot, and so on.

This relation of belonging is denoted by the symbol $\in$. So the fact that an element *a belongs to a set A* is denoted by $a \in A$. This is usually said "a is an element of A." If an object *b does not belong to A*, then we will write $b \notin A$. It is important to realize that for each object a and for each set A *we can have only one of two possible cases*, namely that $a \in A$ or $a \notin A$.

For example, if we define the set $\mathbb{N}$ to be the set of all counting (or natural) numbers, then we observe that $2 \in \mathbb{N}, 3 \in \mathbb{N}, 1034 \in \mathbb{N}$, but $-2 \notin \mathbb{N}, \frac{1}{3} \notin \mathbb{N}$, $\sqrt[3]{12} \notin \mathbb{N}$, and so on.

For a finite set A we can *list all its elements* (this is one way of defining a set). If the elements of A are denoted by $a_1, a_2, \ldots, a_n$ (here the $\ldots$ indicates that the pattern continues), then we write A in the following standard form,

$$A = \{a_1, a_2, \ldots, a_n\}.$$

For example, $A = \{1, 3, 5, 10\}$ means that the set A consists of the numbers $1, 3, 5, 10$. In such a case it is easy to see if an object belongs to this set or not. For instance, the number 1 is an element of this set, while the number 11 is not.

For **another example** let $B = \{\triangleright, \subset, \subseteq, \trianglelefteq\}$. The symbols $\triangleright, \subset, \subseteq, \trianglelefteq$ are elements of this set, but $\triangleleft$ is not which means $\triangleleft \notin B$.

We note that the element a and the set $\{a\}$ are different entities; here $\{a\}$ is a set, having only one element a (sometimes called *a singleton*). Thus the presence or absence of { and } is very important.

However, even when a set only has a finite number of elements, it is sometimes not easy to define the set by just listing its elements. The set could be very large, as is the case when we consider the set of all atoms in our pencil.

In this case, we can *assign a certain property that uniquely characterizes elements and unifies them within the given set.* This is a common way of defining a set. If $P(x)$ is some defining property that an element x of a set A either has or does not have then we use the notation

$$A = \{x \mid P(x)\}.$$

This is literally described as "the set of x such that $P(x)$." Some authors use the notation $\{x : P(x)\}$ instead.

For example, the set of all real numbers belonging to the segment $[2,5]$ is written as $\{x \mid x \in \mathbb{R} \text{ and } 2 \leq x \leq 5\}$ or as $\{x \in \mathbb{R} | 2 \leq x \leq 5\}$. Here $\mathbb{R}$ is the set of all real numbers.

It is important to note that the same set can be determined by distinct defining properties. **For example**, the set X of all solutions of the equation $x^2 - 3x + 2 = 0$, and the set Y consisting of the first two counting numbers have the same elements, namely, the numbers 1 and 2.

We use the following conventional notation for the following sets of numbers.

$\mathbb{N}$ *is the set of all natural numbers, so* $\mathbb{N} = \{1,2,3,\ldots\}$*;*
$\mathbb{Z}$ *is the set of all integers, so* $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$*;*
$\mathbb{Q}$ *is the set of all rational numbers, so* $\mathbb{Q} = \{a/b | a, b \in \mathbb{Z}, b \neq 0\}$*;*
$\mathbb{R}$ *is the set of all real numbers.*

By common agreement, the number 0 *is not a natural number*. We write $\mathbb{N}_0$ for the set consisting of all natural numbers and the number 0 (*the set of whole numbers*).

Now we shall introduce the most important concepts related to sets.

Definition 1.1.1. *Two sets A and B are called equal if every element of A is an element of B and conversely, every element of B is an element of A. We then write A = B.*

A very important set is *the empty set*.

Definition 1.1.2. *A set is said to be empty if it has no elements. The empty set is denoted by* $\emptyset$.

Definition 1.1.1 shows that the empty set is unique. The empty set is always obtained if there is a contradictory property. For example, $\emptyset = \{x \mid x \in \mathbb{R} \text{ and } 2^x < 0\}$.

Definition 1.1.3. *A set A is a subset of a set B if every element of A is an element of B. This is denoted by* $A \subseteq B$.

Note that the sets *A and B are equal if and only if* $A \subseteq B$ *and* $B \subseteq A$. Indeed, in this case, every element of A is an element of B, and every element of B is an element of A. From this definition we see that the empty set is a subset of each set, and every set A is a subset of itself.

Definition 1.1.4. *A subset A of a set B is called a proper subset of B if A is a subset of B and* $A \neq B$. *This is written* $A \subset B$ *or* $A \subsetneqq B$.

In this case there exists an element $x \in A$ such that $x \notin B$. So the only subset of a nonempty set A that is not a proper subset of A is the set A itself. All other subsets of A are proper subsets of A.

Example. Let A be the set of all rectangles in the plane. Then the set B of all squares in the plane is a proper subset of A.

Again we emphasize the notation. If A is a set and a is an element of A then it is correct to write $a \in A$, but in general it will not be true that $\{a\} \in A$. However $a \in A$ if and only if $\{a\} \subseteq A$. For **example,** let A be the set of all subsets of the set $B = \{\triangleright, \subset, \subseteq\}$. Then $\{\triangleright\} \in A$, $\{\triangleright\} \subseteq B$, but of course, $\{\triangleright\} \notin B$.

Definition 1.1.5. *Let A be a set. Then the set of all subsets of A is denoted by* $\mathfrak{B}(A)$ *and is called the Boolean, or power set, of A. Thus* $\mathfrak{B}(A) = \{X \mid X \subseteq A\}$.

Example. Let $B = \{\triangleright, \subseteq, \trianglelefteq\}$. In this case

$$\mathfrak{B}(B) = \{\emptyset, \{\triangleright\}, \{\subseteq\}, \{\trianglelefteq\}, \{\triangleright, \subseteq\}, \{\triangleright, \trianglelefteq\}, \{\subseteq, \trianglelefteq\}, \{\triangleright, \subseteq, \trianglelefteq\}\}$$

is the power set of B. Notice that the set B consists of three elements, while the set $\mathfrak{B}(B)$ consists of eight elements, and that $2^3 = 8$. This is not a coincidence, but illustrates the general rule stating that if a set consists of n elements, then its power set consists of 2^n elements. This rule plays an important role in set theory and can be extended to the infinite case.

Next we introduce some operations on sets.

Definition 1.1.6. *Let A and B be sets. Then* $A \cap B$ *is the set of all elements that belong to A and to B simultaneously. This is called the intersection of A and B. Thus*

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

Example. If $A = \{1, 2, 3, 4, 5\}, B = \{3, 5, 6, 10\}$, then $A \cap B = \{3, 5\}$.

Definition 1.1.7. *Let A and B be sets. Then $A \cup B$ is the set of all elements that belong to A or to B, or both, called the union of A and B. Thus*

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

Example. If $A = \{1,2,3,4,5\}, B = \{3,5,6,10\}$, then $A \cup B = \{1,2,3,4,5,6,10\}$.

Definition 1.1.8. *Let A and B be sets. Then $A \setminus B$ is the set of all elements that belong to A but not to B, called the difference of A and B. Thus*

$$A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}.$$

If $B \subseteq A$, then $A \setminus B$ is called the complement of B in A.

Example. Let A be the set of all right-handed people, and let B be the set of all people with brown hair.

Then:

$A \cap B$ is the set of all right-handed, brown-haired people,

$A \cup B$ is the set of all people who are right-handed or brown-haired or both,

$A \setminus B$ is the set of all people who are right-handed but not brown-haired, and

$B \setminus A$ is the set of all people who have brown hair but are not right-handed.

Example. The set of irrational numbers is the complement of the set $\mathbb{Q}$ of rational numbers in the set $\mathbb{R}$ of all real numbers.

The set $\{0\}$ is the complement of the set $\mathbb{N}$ of all natural numbers in the set $\mathbb{N}_0$ of whole numbers.

We collect together some of the standard results concerning operations on sets.

Theorem 1.1.9. *Let A, B, and C be sets.*

(i) *$A \subseteq B$ if and only if $A \cap B = A$ or $A \cup B = B$. In particular, $A \cup A = A = A \cap A$ (the idempotency of intersection and union).*

(ii) *$A \cap B = B \cap A$ and $A \cup B = B \cup A$ (the commutative property of intersection and union).*

(iii) *$A \cap (B \cap C) = (A \cap B) \cap C$ and $A \cup (B \cup C) = (A \cup B) \cup C$ (the associative property of intersection and union).*

(iv) *$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (the distributive property).*

(v) $A \setminus (A \setminus B) = A \cap B$.

(vi) $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$.

(vii) $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$.

Proof. The proofs of the majority of these assertions are easy to write using the definitions. However, to indicate how the proofs may be written, we give a proof of (iv).

Let $x \in A \cap (B \cup C)$. It follows from the definition that $x \in A$ and $x \in B \cup C$. Since $x \in B \cup C$ either $x \in B$ or $x \in C$ and hence either x is an element of both sets A and B, or x is an element of both sets A and C. Thus $x \in A \cap B$ or $x \in A \cap C$, which is to say that $x \in (A \cap B) \cup (A \cap C)$. This shows that $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$.

Conversely, since $B \subseteq B \cup C$ we have $A \cap B \subseteq A \cap (B \cup C)$. Likewise $A \cap C \subseteq A \cap (B \cup C)$ and hence $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$.

We can extend the notions of intersection and union to arbitrary families of sets. Let $\mathfrak{S}$ be a family of sets. Thus the elements of $\mathfrak{S}$ are also sets.

Definition 1.1.10. *The intersection of the family $\mathfrak{S}$ is the set of elements that belong to each set S from the family $\mathfrak{S}$ and is denoted by $\cap\mathfrak{S}$. Thus:*

$$\cap\mathfrak{S} = \bigcap_{S \in \mathfrak{S}} S = \{x \mid x \in S \text{ for each set } S \in \mathfrak{S}\}.$$

Definition 1.1.11. *The union of the family $\mathfrak{S}$ is the set of elements that belong to at least one set S from the family $\mathfrak{S}$ and is denoted by $\cup\mathfrak{S}$. Thus:*

$$\cup\mathfrak{S} = \bigcup_{S \in \mathfrak{S}} S = \{x \mid x \in S \text{ for some set } S \in \mathfrak{S}\}.$$

The idea of an ordered pair of real numbers is very familiar to most students of mathematics and we now extend this idea to arbitrary sets A and B. A pair of elements (a,b) where $a \in A, b \in B$, taken in the given order, is called *an ordered pair*. By definition, $(a,b) = (a_1,b_1)$ if and only if $a = a_1$ and $b = b_1$.

Definition 1.1.12. *Let A and B be sets. Then the set $A \times B$ of all ordered pairs (a,b), where $a \in A, b \in B$, is called the Cartesian product of the sets A and B. If $A = B$, then we call $A \times A$ the Cartesian square of the set A and write $A \times A$ as A^2.*

The real plane $\mathbb{R}^2$ is a natural **example** of a Cartesian product. The Cartesian product of two segments of the real number line could be interpreted geometrically as a rectangle whose sides are these segments.

Example. If $A = \{1,2,3,4\}$ and $B = \{a,b,c\}$, the Cartesian product $A \times B = \{(1,a),(1,b),(1,c),(2,a),(2,b),(2,c),(3,a),(3,b),(3,c),(4,a),(4,b),(4,c)\}$.

To make sure that none of the ordered pairs are missed, remember that if the set A consists of four elements, and the set B consists of three elements, their product must have $4 \times 3 = 12$ elements. More generally, if A has m elements and B has n elements, then $A \times B$ has mn elements.

It is easy to extend the notion of a Cartesian product of two sets to the Cartesian product of a finite family of sets.

Definition 1.1.13. *Let n be a natural number and let $A_1,\ldots,A_n$ be sets. Then the set*

$$A_1 \times \cdots \times A_n = \prod_{1\le i\le n} A_i$$

of all ordered n-tuples $(a_1,\ldots,a_n)$ where $a_j \in A_j$, for $1 \le j \le n$, is called the Cartesian product of the sets $A_1,\ldots,A_n$.

Here $(a_1,\ldots,a_n) = (b_1,\ldots,b_n)$ if and only if $a_1 = b_1,\ldots,a_n = b_n$.

The element a_j is called *the j-th coordinate or j-th component of* $(a_1,\ldots,a_n)$.

If $A_1 = \cdots = A_n = A$ we call $\underbrace{A \times A \times \cdots \times A}_{n}$ *the n-th Cartesian power A^n* of the set A.

We shall use the convention that if A is a nonempty set then A^0 will denote a one-element set and we shall denote A^0 by $\{*\}$, where $*$ denotes the unique element of A^0. Naturally, $A^1 = A$.

Example. The most natural example of a Cartesian product of more than two sets is real three-dimensional space $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$.

We note that the commutative law is not valid in general for Cartesian products, which is to say that in general $A \times B \neq B \times A$ if $A \neq B$. The same can also be said for the associative law: It is normally the case that $A \times (B \times C)$, $(A \times B) \times C$, and $A \times B \times C$ are distinct sets.

Exercise Set 1.1

In each of the following questions explain your reasoning by giving a proof of your assertion or by using appropriate examples.

1.1.1. Which of the following assertions are valid for all sets A, B, and C?

(i) If $A \notin B$ and $B \notin C$, then $A \notin C$.

(ii) If $A \notin B$ and $B \nsubseteq C$, then $A \notin C$.

1.1.2. Which of the following assertions are valid for all sets A, B, and C?

(i) If $A \subseteq B$, $A \neq B$ and $B \subseteq C$, then $C \nsubseteq A$.
(ii) If $A \subseteq B$, $A \neq B$ and $B \in C$, then $A \notin C$.

1.1.3. Give examples of sets A, B, C, D satisfying all of the following conditions: $A \subseteq B, A \neq B, B \in C, C \in D$.

1.1.4. Give examples of sets A, B, C satisfying all the following conditions: $A \in B, B \in C$, but $A \notin C$.

1.1.5. Let

$A = \{x \in \mathbb{Z} | x = 2y \text{ for some } y \geq 0\}$;
$B = \{x \in \mathbb{Z} | x = 2y - 1 \text{ for some } y \geq 0\}$;
$C = \{x \in \mathbb{Z} |\ x < 10\}$.
Find $\mathbb{Z} \backslash A$ and $\mathbb{Z} \backslash (A \cap B)$

1.1.6. Let

$A = \{x \in \mathbb{Z} | x = 2y \text{ for some } y \geq 0\}$;
$B = \{x \in \mathbb{Z} | x = 2y - 1 \text{ for some } y \geq 0\}$;
$C = \{x \in \mathbb{Z} |\ x < 10\}$.
Find $\mathbb{Z} \backslash C$ and $C \backslash (A \cup B)$.

1.1.7. Let S be the set of all complex roots of the polynomial $f(X) \in \mathbb{R}[X]$. Suppose that $f(X) = g(X)h(X)$. Let S_1 (respectively S_2) be the set of all roots of the polynomial $g(X)$ (respectively $h(X)$). Prove that $S = S_1 \cup S_2$.

1.1.8. Let $g(X)$ and $h(X)$ be polynomials with real coefficients. Let S_1 (respectively S_2) be the set of all real roots of the polynomial $g(X)$ (respectively $h(X)$). Let S be the set of all real roots of the polynomial $f(X) = (g(X))^2 + (h(X))^2$. Prove that $S = S_1 \cap S_2$.

1.1.9. Prove that $\mathfrak{B}(A \cap B) = \mathfrak{B}(A) \cap \mathfrak{B}(B)$.

1.1.10. Prove that the equation $\mathfrak{B}(A) \cup \mathfrak{B}(B) = \mathfrak{B}(A \cup B)$ implies that either $A \subseteq B$ or $B \subseteq A$.

1.1.11. Prove that if A, B are sets then $A \setminus (A \setminus B) = A \cap B$.

1.1.12. Prove that if A, B, C are sets then $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$.

1.1.13. Let $A_n = [0, 1/n)$, for each natural number n. What is $\cap_{n \geq 1} A_n$?

1.1.14. Let $A_n = (0, 1/n]$, for each natural number n. What is $\cap_{n \geq 1} A_n$?

1.1.15. Do there exist nonempty sets A, B, C such that $A \cap B \neq \emptyset, A \cap C = \emptyset, (A \cap B) \backslash C = \emptyset$?

1.1.16. Let $A = \{1,2,3,4,5,6,7\}, B = \{2,5,7,8,9,10\}$. Find $A \cap B, A \cup B, A \setminus B, B \setminus A$, the complement of A in $\mathbb{N}$, the number of elements in $A \times B$, and the number of elements in $\mathfrak{B}(A)$.

1.1.17. Let A, B, C be sets. Prove or disprove: $(A \cap B) \times C = (A \times C) \cap (B \times C)$.

1.1.18. Let A, B, C be sets. Prove or disprove: $(A \cup B) \setminus C = (A \setminus C) \cup (B \setminus C)$.

1.1.19. The symmetric difference of two sets A, B is defined by $A \triangle B = (A \cup B) \setminus (A \cap B)$. Prove that $A \triangle B = (A \setminus B) \cup (B \setminus A)$. Also prove that $A \cap (B \triangle C) = (A \cap B) \triangle (A \cap C)$ and $A \triangle (A \triangle B) = B$.

1.1.20. Is it possible to find three sets A, B, C such that $A \cap B \neq \emptyset, A \cap C \neq \emptyset, B \cap C \neq \emptyset$, but $A \cap B \cap C = \emptyset$.

1.2 SET MAPPINGS

The notion of a mapping (or function) plays a key role in mathematics.

A *mapping (or a function)* f from a set A to a set B is defined if for each element of A there is a rule that associates a uniquely determined element of B. This is usually written $f : A \longrightarrow B$. If $a \in A$, then the unique element $b \in B$, which corresponds to a, is denoted by $f(a)$ and we sometimes write $a \longmapsto b$. We say that $b = f(a)$ is an *image* of a, and a is a *preimage or inverse image* of b. Each element of A has one and only one image. However, an element $b \in B$ can have several preimages or no preimage at all. If $b \in B$, then we denote the set of preimages of b by $f^{-1}(b) = \{a \in A | f(a) = b\}$. Of course $f^{-1}(b) = \emptyset$ if there are no preimages of b.

The set A is called the *domain* of the mapping f, while the set of all images of all elements of A is a subset of B called the *range* of f which we denote by $\mathbf{Im}(f)$. The set B is usually called the *codomain* of f.

Example. Functions should look familiar to you since you already worked with them in Calculus courses. You can look at the function $y = x^2$ defined on the set $\mathbb{R}$ of real numbers as a mapping of the set $\mathbb{R}$ of real numbers to itself. Here the law of association is the unique number $y = x^2$, corresponding to each $x \in \mathbb{R}$. In this case the domain is $A = \mathbb{R}$, and the range is $B = \mathbf{Im}(f) = \mathbb{R}^+$ the set of all nonnegative real numbers.

We consider a further example having no relation to Calculus.

Example. Let A be the set of all people, B (respectively C) be the set of all males (respectively all females). Define the function $m : A \longrightarrow B$ (respectively $w : A \longrightarrow C$) by the rule that, for each person a, the image $m(a)$ is his/her father (respectively $w(a)$ is his/her mother).

A function can be thought of as a *correspondence* between sets A and B and in particular as a set of ordered pairs (a,b) where the first element a of the pair belongs to the first set A, the domain, and the second element b of the pair belongs to the second set B, the corresponding codomain. Note, that in terms of the Cartesian product, a correspondence between A and B is a subset of $A \times B$.

If $A = B$, then we will say there is a correspondence or a *relation* (or more precisely a *binary relation*) between the elements of A.

Example. Define the mapping $f : A \longrightarrow B$, where $A = \{-1,2,3,5\}$, $B = \{0,2,8,9,11\}$ by the rule: $-1 \longmapsto 2$, $2 \longmapsto 0$, $3 \longmapsto 9$, $5 \longmapsto 8$. The mapping f is defined as the set $\{(-1,2),(2,0),(3,9),(5,8)\}$ and we write $f(-1) = 2, f(2) = 0$ and so on.

As you can see, not all elements from B are involved: $11 \in B$ does not have a match in A. Thus $f^{-1}(11) = \emptyset$.

Thus, with each function $f : A \longrightarrow B$ we can form the set $\{(x,f(x)) \mid x \in A\} \subseteq A \times B$. This subset is called *the graph* of the function f.

Note that not every correspondence can serve as the graph of a function. Only a set of ordered pairs in which each element of the domain has only *one* element associated with it in the range is the graph of a function.

For example, the set of ordered pairs $\{(-1,2),(2,0),(3,9),(5,8),(-1,11)\}$ defines a correspondence between the sets $A = \{-1,2,3,5\}$ and $B = \{0,2,8,9,11\}$ but this does not correspond to a function since the element $-1 \in A$ is connected with two elements, 11 and 2, of B.

For further **examples**, let Φ denote the relation on the set of all people where $(a,b) \in \Phi$ means that a and b are people who have cars of the same brand (let's say Mercedes). This relation will not be a function.

Next, let A be the set of all points in a plane, and let B be the set of all lines in this plane. Let Γ be the correspondence defined by $(a,b) \in \Gamma$, if the point a belongs to the line b. Again this is not a function, since a given point will lie on many lines.

Definition 1.2.1. *The functions $f : A \longrightarrow B$ and $g : C \longrightarrow D$ are said to be equal if $A = C, B = D$ and $f(a) = g(a)$ for each element $a \in A$.*

Some useful and common terminology can be found in the following definition.

Definition 1.2.2. *Let $f : A \longrightarrow B$ be a function.*

(i) *The function f is said to be injective (or one-to-one) if every pair of distinct elements of A have distinct images.*

(ii) *The function f is said to be surjective (or onto) if* $\mathbf{Im} f = B$.
(iii) *The function f is said to be bijective if it is injective and surjective. In this case f is a one-to-one, onto correspondence.*

Examples. First, observe that the function $f : \mathbb{R} \longrightarrow \mathbb{R}$, satisfying $f(x) = x^2$ for all $x \in \mathbb{R}$, is neither injective nor surjective, since, for example, $f(-2) = f(2) = 4$ and $-1 \neq x^2$, for all $x \in \mathbb{R}$. However, the function $f_1 : \mathbb{R}^+ \longrightarrow \mathbb{R}$, satisfying $f_1(x) = x^2$ for all $x \in \mathbb{R}^+$, is injective and the function $f_2 : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$, satisfying $f_2(x) = x^2$ for all $x \in \mathbb{R}^+$, is injective and surjective (so is bijective).

Another familiar example is the function $y = \ln x$. Here we have a bijective mapping from the set $\mathbb{R}_+ \backslash \{0\}$ of all positive real numbers to the set $\mathbb{R}$ of all real numbers.

The correspondence $\{(-1,2),(2,0),(3,9),(5,8)\}$ considered above, from the set $A = \{-1,2,3,5\}$ into the set $B = \{0,2,8,9,11\}$ is injective, but will only be bijective if we delete 11 from the set B.

The function $m : A \longrightarrow B$ from the set A of all people to the set B of all males, where $m(a)$ is the father of person a, is not injective and not surjective.

The following statement is immediate from the definitions.

Proposition 1.2.3. *Let $f : A \longrightarrow B$ be a function. Then*

(i) *f is injective if and only if every element of B has at most one preimage;*
(ii) *f is surjective if and only if every element of B has at least one preimage;*
(iii) *f is bijective if and only if every element of B has exactly one preimage.*

We say that a set A *is finite* if there is a positive integer n, for which there exists a bijective mapping $A \longrightarrow \{1,2,\ldots,n\}$. Thus we can count the elements of A and the positive integer n is called the order of the set A; we will write this as $|A| = n$ or $\mathbf{Card}\ A = n$. The empty set is finite and its order is 0. A set that is not finite is called *infinite*. The following assertions are also easy to see.

Corollary 1.2.4. *Let A and B be finite sets and let $f : A \longrightarrow B$ be a mapping.*

(i) *If f is injective, then $|A| \leq |B|$;*
(ii) *If f is surjective, then $|A| \geq |B|$;*
(iii) *If f is bijective, then $|A| = |B|$.*

The next result is more interesting.

Corollary 1.2.5. *Let A be a finite set and let $f : A \longrightarrow A$ be a mapping.*

(i) *If f is injective, then f is bijective;*
(ii) *If f is surjective, then f is bijective.*

Proof.

(i) If f is injective it is evident from Corollary 1.2.4 that $|A| \leq |\mathbf{Im}(f)|$. Since also $|\mathbf{Im}(f)| \leq |A|$, it follows that $|A| = |\mathbf{Im}(f)|$. Since $\mathbf{Im}(f)$ is a subset of A, this equation implies $A = \mathbf{Im}(f)$, so f is surjective and hence bijective.
(ii) Now suppose that f is surjective. If $a, b \in A$, $a \neq b$ and $f(a) = f(b)$ then the map $g : A \setminus \{a\} \longrightarrow A$ is also surjective and by Corollary 1.2.4 it follows that $|A \setminus \{a\}| \geq |A|$, which is a contradiction. Hence $a = b$ and f is injective.

Definition 1.2.6. *Let A be a set. The mapping $\varepsilon_A : A \longrightarrow A$, defined by $\varepsilon_A(a) = a$, for each $a \in A$, is called the identity mapping of A.*

If C is a subset of A, then the mapping $j_C : C \longrightarrow A$, defined by $j_C(c) = c$ for each element $c \in C$, is called a canonical injection or an identical embedding.

Definition 1.2.7. *Let $f : A \longrightarrow B$ and $g : C \longrightarrow D$ be mappings. Then we say that f is the restriction of g, or g is an extension of f, if $A \subseteq C, B \subseteq D$ and $f(a) = g(a)$ for each element $a \in A$.*

For example, a canonical injection is the restriction of the corresponding identity mapping. Note that a restriction of g is uniquely defined once the subsets A and B have been specified; however there are many different extensions of a mapping. We introduce our next topic rather informally.

Definition 1.2.8. *A set A is called countable if there exists a bijective mapping $f : \mathbb{N} \longrightarrow A$.*

In the case when A is countable, we often write $a_n = f(n)$ for each $n \in \mathbb{N}$. Then

$$A = \{a_1, a_2, \ldots, a_n, \ldots\} = \{a_n \mid n \in \mathbb{N}\}.$$

Thus the elements of a countable set can be indexed (or numbered) by the set of all positive integers. Conversely, if all elements of an infinite set A can be indexed using natural numbers, then A is countable. The bijection from $\mathbb{N}$ to A here is natural: every natural number n corresponds to the element a_n of A with index n.

Proposition 1.2.9. *Let A and B be sets and suppose that A is countable. If there exists a bijective mapping $f : B \longrightarrow A$ (respectively $g : A \longrightarrow B$), then B is countable.*

Proof. Since A is countable, we can write $A = \{a_n | n \in \mathbb{N}\}$.

First suppose that there is a bijection $g : A \longrightarrow B$. Consider the mapping $g_1 : \mathbb{N} \longrightarrow B$, defined by $g_1(n) = g(a_n), n \in \mathbb{N}$. It is easy to see that g_1 is bijective.

Suppose now that there exists a bijection $f : B \longrightarrow A$. Then each element a of A has exactly one preimage $f^{-1}(a)$. Consider the mapping $f_1 : \mathbb{N} \longrightarrow B$ defined by $f_1(n) = f^{-1}(a_n), n \in \mathbb{N}$. It is easy to see that f_1 is bijective.

Theorem 1.2.10.

(i) *Let A be an infinite set. Then A contains a countable subset;*

(ii) *Let A be a countable set and let B be a subset of A. If B is infinite, then B is countable;*

Proof.

(i) Since A is infinite it is not empty so choose $a_1 \in A$. The subset $A \setminus \{a_1\}$ is also not empty, therefore we can choose an element a_2 in this subset. Since A is infinite, $A \setminus \{a_1, a_2\} \neq \emptyset$, so that we can choose an element a_3 in this subset and so on. This process cannot terminate after finitely many steps because A is infinite. Hence A contains the infinite subset $\{a_n \mid n \in \mathbb{N}\}$, which is countable.

(ii) Let $A = \{a_n \mid n \in \mathbb{N}\}$. Then there is a least positive integer $k(1)$ such that $a_{k(1)} \in B$ and we put $b_1 = a_{k(1)}$. There is a least positive integer $k(2)$ such that $a_{k(2)} \in B \setminus \{b_1\}$. Put $b_2 = a_{k(2)}$, and so on. This process cannot terminate since B is infinite. Then all the elements of B will be indexed by positive integers.

Notice that Theorem 1.2.10 implies that a subset of a countable set is either countable or finite.

Corollary 1.2.11. *Every infinite subset of $\mathbb{N}$ is countable.*

Corollary 1.2.12. *Let A and B be sets. If A is countable and there is an injective mapping $f : B \longrightarrow A$, then B is finite or countable.*

Proof. We consider the mapping $f_1 : B \longrightarrow \mathbf{Im}f$ defined by $f_1(b) = f(b)$, for each element $b \in B$. By this choice, f_1 is surjective. Since f is injective, f_1 is

also injective and hence f_1 is bijective. Finally, Theorem 1.2.10 implies that $\mathbf{Im}f$ is finite or countable.

Example. The set of integers is countable. We can construct a bijective mapping $f : \mathbb{N} \longrightarrow \mathbb{Z}$, informally, in the following way

$$\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & \dots \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \dots . \\ 0 & 1 & -1 & 2 & -2 & 3 & -3 & \dots \end{array}$$

It is a little bit trickier to find a bijective mapping between $\mathbb{N}$ and the set of all rational numbers $\mathbb{Q}$ and we here indicate informally how to do this. Put $\mathbb{Q}_n = \{\frac{m}{k} \mid |m| + |k| = n\}$, for each $n \in \mathbb{N}$. Then each subset $\mathbb{Q}_n$ is finite and the elements of $\mathbb{Q}_n$ can be ordered in their natural order. Now we construct a bijective mapping $r : \mathbb{N} \longrightarrow \mathbb{Q}$. We have $\mathbb{Q}_1 = \{\frac{0}{\pm 1} = 0\}$, so put $r(1) = 0$. Further, $\mathbb{Q}_2 = \{-1 = \frac{-1}{1} = \frac{1}{-1}, \frac{0}{\pm 2} = 0, 1 = \frac{1}{1} = \frac{-1}{-1}\}$. Since 0 already has a preimage, put $r(2) = -1$, $r(3) = 1$. For the next step we consider $\mathbb{Q}_3 = \{-2 = \frac{-2}{1} = \frac{2}{-1}, \frac{-1}{2} = \frac{-1}{2} = \frac{1}{-2}, \frac{0}{\pm 3} = 0, \frac{1}{2} = \frac{1}{2} = \frac{-1}{-2}, 2 = \frac{2}{1} = \frac{-2}{-1}\}$.

Again 0 already has a preimage, so put $r(4) = -2, r(5) = \frac{-1}{2}, r(6) = \frac{1}{2}, r(7) = 2$. Consider next $\mathbb{Q}_4 = \{-3 = \frac{-3}{1} = \frac{3}{-1}, \frac{-1}{3} = \frac{-1}{3} = \frac{1}{-3}, -1 = \frac{-2}{2} = \frac{2}{-2}, 0 = \frac{0}{\pm 4}, \frac{1}{3} = \frac{-1}{-3}, 1 = \frac{2}{2} = \frac{-2}{-2}, 3 = \frac{3}{1} = \frac{-3}{-1}\}$. The numbers $0, -1, 1$ have preimages, thus we need to index the numbers $-3, -\frac{1}{3}, \frac{1}{3}, 3$, so put $r(8) = -3, r(9) = \frac{-1}{3}, r(10) = \frac{1}{3}, r(11) = 3$. If we continue this process we will index all rational numbers using natural numbers.

An important natural question arises: Is there an infinite set that is not countable? The answer to this question is yes and was obtained by Georg Cantor who proved that *the set* $[0, 1]$*, and therefore the set of all real numbers, is not countable.* We shall not pursue this topic further here.

As we saw here, to establish that two sets have the same number of elements there is no need to count these elements. It is sufficient to establish the existence of a bijective mapping between these sets. This idea is really at the heart of the abstract notion of a number. By extending this to arbitrary sets we arrive at the concept of the *cardinality* of a set.

Definition 1.2.13. *Two sets A and B are called equipollent, if there exists a bijective mapping* $f : A \longrightarrow B$.

We will denote this by $|A| = |B|$.

If A and B are finite sets, then A and B are equipollent precisely when these sets have the same number of elements. More generally when two sets are equipollent we say that they have the same cardinal number. This allows us to establish an ordering of the set of cardinal numbers, but we refrain from pursuing this topic.

Exercise Set 1.2

In each of the following questions explain your reasoning, either by giving a proof of your assertion or a counterexample.

1.2.1. Let $\Phi = \{(x,y) \in \mathbb{N} \times \mathbb{N} \mid 3x = y\}$. Is Φ a function?

1.2.2. Let $\Phi = \{(x,y) \in \mathbb{N} \times \mathbb{N} \mid 3x = 5y\}$. Is Φ a function?

1.2.3. Let $\Phi = \{(x,y) \in \mathbb{N} \times \mathbb{N} \mid x^2 = y^2\}$. Is Φ a function?

1.2.4. Let $\Phi = \{(x,y) \in \mathbb{N} \times \mathbb{N} \mid x = y^4\}$. Is Φ a function?

1.2.5. Let $f : \mathbb{Z} \longrightarrow \mathbb{N}_0$ be the mapping defined by $f(n) = |n|$, where $n \in \mathbb{Z}$. Is f injective? Is f surjective?

1.2.6. Let $f : \mathbb{N} \longrightarrow \{x \in \mathbb{Q} \mid x > 0\}$ be the mapping defined by $f(n) = \frac{n}{n+1}$, where $n \in \mathbb{N}$. Is f injective? Is f surjective?

1.2.7. Let $f : \mathbb{N} \longrightarrow \mathbb{N}$ be the mapping, defined by $f(n) = (n+1)^2$, where $n \in \mathbb{N}$. Is f injective? Is f surjective?

1.2.8. Let $f : \mathbb{N} \longrightarrow \mathbb{N}$ be the mapping, defined by $f(n) = \frac{n^2+n}{2}$, where $n \in \mathbb{N}$. Is f injective? Is f surjective?

1.2.9. Let $f : \mathbb{Z} \longrightarrow \mathbb{Z} \times \mathbb{Z}$ be the mapping, defined by the rule $f(n) = (n+1, n)$, where $n \in \mathbb{Z}$. Is f injective? Is f surjective?

1.2.10. Let $f : \mathbb{Z} \longrightarrow \mathbb{Z} \times \mathbb{Z}$ be the mapping, defined by $f(n) = (n, n^4)$, where $n \in \mathbb{Z}$. Is f injective? Is f surjective?

1.2.11. Let $f : \mathbb{Q} \longrightarrow \mathbb{R}$ be the map defined by $f(a) = a$, for all $a \in \mathbb{Q}$. Define $g_1 : \mathbb{R} \longrightarrow \mathbb{R}$ by $g_1(a) = a$, for all $a \in \mathbb{R}$ and define $g_2 : \mathbb{R} \longrightarrow \mathbb{R}$ by

$$g_2(a) = \begin{cases} a, \text{ if } a \in \mathbb{Q} \\ 1, \text{ if } a \notin \mathbb{Q} \end{cases}$$

Show that g_1, g_2 are both extensions of f.

1.2.12. Let A and B be finite sets, with $|A| = a, |B| = b$. Find the number of injective mappings from A to B.

1.2.13. Let $f : A \longrightarrow B$ be a function from the set A to the set B and let U, V be subsets of A. Give a proof or counterexample to the statement: $f(U \cap V) = f(U) \cap f(V)$.

1.2.14. Let A be a finite set. Prove that if $f : A \longrightarrow A$ is an injective function then f is also surjective.

1.2.15. Let $f : A \longrightarrow B$ be a function from the set A to the set B and let U, V be subsets of B. Give a proof or counterexample to the statement: $f^{-1}(U) \cap f^{-1}(V) = f^{-1}(U \cap V)$.

1.2.16. Let $f : \mathbb{N}_0 \longrightarrow \mathbb{N}_0$ be the mapping defined by $f(n) = n^2 + 3n$, where $n \in \mathbb{N}_0$. Is f injective? Is f surjective?

1.2.17. Let $f : A \longrightarrow B$ be a function from the set A to the set B and let U, V be subsets of B. Give a proof or counterexample to the statement: $f^{-1}(U) \cup f^{-1}(V) = f^{-1}(U \cup V)$.

1.2.18. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a bijection. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the map defined by $g(x) = f(5x+3)$. Is g injective? Is g surjective?

1.2.19. Prove that a countable union of countable sets is again countable.

1.2.20. Prove that if A and B are countable sets then $A \times B$ is also countable.

1.3 PRODUCTS OF MAPPINGS AND PERMUTATIONS

We next consider the product of two mappings. This product will allow us to construct new mappings based on given ones, but it is not defined in all cases. If $f : A \longrightarrow B$ and $g : C \longrightarrow D$ are mappings, then the product of g and f is defined only when $B = C$.

Definition 1.3.1. *Let $f : A \longrightarrow B$ and $g : B \longrightarrow C$ be mappings. The mapping $g \circ f : A \longrightarrow C$, defined by the rule*

$$g \circ f(a) = g((f(a)) \text{ for each } a \in A$$

is called the product or the composite of g and f.

We think of this as follows. First the mapping f acts on the element $a \in A$, and then the mapping g acts on the image $f(a)$ (the result of the first mapping f applied to a). Thus, when we write $g \circ f$, the mapping f is done first, opposite to the usual rules for reading. There will be one important exception to this general rule, which we will discuss later.

Example. We are familiar with composition of real functions, so there are many standard examples that can be used to illustrate the product of functions. For example, let $f : \mathbb{R} \longrightarrow \mathbb{R}$ and $g : \mathbb{R} \longrightarrow \mathbb{R}$ be defined by $f(x) = 4x - 1$, $g(x) = x^2 + 1$, for all $x \in \mathbb{R}$. In this case, both products $g \circ f$ and $f \circ g$ are defined and we now evaluate these compositions, using two slightly different methods. For every element $x \in \mathbb{R}$ we have

$$(g \circ f)(x) = g(f(x)) = f(x)^2 + 1 = (4x-1)^2 + 1 = 16x^2 - 8x + 2, \text{ and}$$
$$(f \circ g)(x) = f(g(x)) = f(x^2+1) = 4(x^2+1) - 1 = 4x^2 + 3$$

We next consider an example of the product of two nonnumeric functions. Let A be the set of all people and consider the functions $m : A \longrightarrow A$ and $w : A \longrightarrow A$, where $m(a)$ is the father of person a, and $w(a)$ is the mother of person a. In this case, both products $m \circ w$ and $w \circ m$ are defined. Then $(m \circ w)(a) = m(w(a))$ is the father of the mother of the person a, which is the grandfather on the mother's side, while $(w \circ m)(a) = w(m(a))$ is the mother of the father of the person a, which is the grandmother on the father's side.

These examples show that the product of mappings is not a commutative operation. In general, the situation when $g \circ f = f \circ g$ is fairly rare. We say that the mappings $f : A \longrightarrow B$ and $g : C \longrightarrow D$ *permute or commute* if both products $g \circ f$ and $f \circ g$ exist (i.e., $C = B$ and $D = A$) and $g \circ f = f \circ g$, in which case $A = B = C = D$.

However, function composition always satisfies the associative property, at least when the products are defined.

Theorem 1.3.2. *Let $f : A \longrightarrow B, g : B \longrightarrow C$ and $h : C \longrightarrow D$ be functions. Then $h \circ (g \circ f) = (h \circ g) \circ f$.*

Proof. We have $g \circ f : A \longrightarrow C, h \circ g : B \longrightarrow D$ and so

$$h \circ (g \circ f) : A \longrightarrow D, (h \circ g) \circ f : A \longrightarrow D.$$

If a is an arbitrary element of A, then

$$(h \circ (g \circ f))(a) = h((g \circ f)(a)) = h(g(f(a))),$$

whereas

$$((h \circ g) \circ f)(a) = (h \circ g)(f(a)) = h(g(f(a))).$$

Hence $\big(h \circ (g \circ f)\big)(a) = \big((h \circ g) \circ f\big)(a)$ for all $a \in A$ which proves that $(h \circ g) \circ f = h \circ (g \circ f)$.

Let $f : A \longrightarrow B$ be a mapping. It is not hard to see that

$$\varepsilon_B \circ f = f \circ \varepsilon_A = f,$$

so the mappings ε_B and ε_A play the role of "left identity" and "right identity" elements, respectively, for the operation of multiplication of mappings. Also it should be noted that there is no "universal" identity element for all mappings.

Definition 1.3.3. *Let $f : A \longrightarrow B$ be a mapping. Then the mapping $g : B \longrightarrow A$ is called an inverse of f if $g \circ f = \varepsilon_A$ and $f \circ g = \varepsilon_B$.*

We remark first that if f has an inverse then it is unique. To show this let $g : B \longrightarrow A$ and $h : B \longrightarrow A$ be mappings satisfying

$$g \circ f = \varepsilon_A, f \circ g = \varepsilon_B \text{ and } h \circ f = \varepsilon_A, f \circ h = \varepsilon_B.$$

Now consider the product $g \circ f \circ h$. We have

$$h = \varepsilon_A \circ h = (g \circ f) \circ h = g \circ (f \circ h) = g \circ \varepsilon_B = g.$$

Theorem 1.3.4. *Let $f : A \longrightarrow B, g : B \longrightarrow A$ be mappings. If $g \circ f = \varepsilon_A$, then f is an injective mapping and g is a surjective mapping.*

Proof. Suppose that A has elements a and c such that $f(a) = f(c)$. Then

$$a = \varepsilon_A(a) = (g \circ f)(a) = g(f(a)) = g(f(c)) = g \circ f(c) = \varepsilon_A(c) = c,$$

which shows that f is injective.

Next, let u be an arbitrary element of A. Then

$$u = \varepsilon_A(u) = g \circ f(u) = g(f(u)),$$

and, in particular, $f(u)$ is a preimage of the element u relative to g. It follows that $\mathbf{Im}\ g = A$, so g is surjective.

Corollary 1.3.5. *Let $f : A \longrightarrow B$ be a mapping. Then f has an inverse mapping if and only if f is bijective. In this case, the inverse mapping is also bijective.*

Proof. Suppose that f has inverse mapping $g : B \longrightarrow A$. Then $g \circ f = \varepsilon_A$ and $f \circ g = \varepsilon_B$. From the first equation and Theorem 1.3.4 it follows that f is injective and g is surjective. Applying Theorem 1.3.4 to the second equation, we deduce that g is injective and f is surjective. It follows that f and g are both bijective.

Conversely, let f be bijective. By Proposition 1.2.3, every element $b \in B$ has exactly one preimage a_b. Thus $f(a_b) = b$ and we may define the mapping $g : B \longrightarrow A$ by $g(b) = a_b$. We show that g is the desired inverse to f. Indeed, if

$b \in B$ then $f(g(b)) = f(a_b) = b$ so $f \circ g = \varepsilon_B$. On the other hand, if $a \in A$ then $f(a)$ has the unique preimage a and so $g(f(a)) = a$ by definition of g. Thus $g \circ f = \varepsilon_A$ and the proof is complete.

Since a bijective mapping f has only one inverse, we use the notation f^{-1} for it. The reader is cautioned that the notation does not mean $1/f(x)$. We note also that by Corollary 1.3.5 the mapping f^{-1} is also bijective. We observe that Corollary 1.3.5 not only proves the existence of the inverse mapping, but also shows how to find it.

For **example**, consider the real functions f and g defined as follows:

$$f(x) = 4x - 1, g(x) = 5x^3 + 1 \text{ for each } x \in \mathbb{R}.$$

Once we know that f, g are bijections we can find their inverses, as usual, by "solving for x in terms of y" and often to see that a function is surjective amounts to doing just that. For example, let $b \in \mathbb{R}$. We find its unique preimage, c, relative to f by solving the equation $b = 4c - 1$. Clearly $c = \frac{b+1}{4}$ and hence $f^{-1} : \mathbb{R} \longrightarrow \mathbb{R}$ is defined as follows:

$$f^{-1}(x) = \frac{x+1}{4}.$$

Of course, we have not shown that f is injective here.

To show that g is injective we have to show that if $5x_1^3 + 1 = 5x_2^3 + 1$ then $x_1 = x_2$; however, this follows since the first equation implies that $x_1^3 = x_2^3$ so $x_1 = x_2$. Furthermore, if $b = 5a^3 + 1$ then we can solve uniquely for a to obtain $a = \sqrt[3]{\frac{b-1}{5}}$ so that the inverse of g is

$$g^{-1}(x) = \sqrt[3]{\frac{x-1}{5}}.$$

We note one further important property of the product of functions that we have already used.

Proposition 1.3.6. *Let $f : A \longrightarrow B, g : B \longrightarrow C$ be mappings.*

(i) *If f and g are injective, then $g \circ f$ is injective;*
(ii) *If f and g are surjective, then $g \circ f$ is surjective;*
(iii) *If f and g are bijective, then $g \circ f$ is bijective.*

This statement can be proved directly using the definitions.

Definition 1.3.7. *Let A be a set. A mapping from A to A is called a transformation of the set A. The set of all transformations of A is denoted by* $\mathbf{P}(A)$ *or* A^A.

We note that a product of two transformation of A is always defined and is again a transformation. Clearly, multiplication of transformations is associative and in this case there exists an identity element, namely the identity transformation ε_A.

Examples. The following transformations play a significant role in geometry. Let a be a line in space and for each point $P \in \mathbb{R}^3$ let Q be the point obtained by rotating P about the line a through an angle α. Thus a acts as the axis of rotation and this defines a transformation of $\mathbb{R}^3$ called the rotation of $\mathbb{R}^3$ about the axis a through angle α.

Another important transformation of the space $\mathbb{R}^3$ is a translation by a given vector. Of course we can consider similar transformations of the plane $\mathbb{R}^2$.

The mapping $f : \mathbb{Z} \longrightarrow \mathbb{Z}$ defined by $f(k) = k^2 + 1$, where $k \in \mathbb{Z}$, is a transformation of the set of integers $\mathbb{Z}$. It is not bijective, therefore it has no inverse. The mapping $g : \mathbb{Z} \longrightarrow \mathbb{Z}$, defined by $g(k) = -k$, where $k \in \mathbb{Z}$ is a bijective transformation of $\mathbb{Z}$, and this transformation is clearly its own inverse.

There are other important transformations. For example, it is well-known that a projection of space onto a plane is an important transformation. This transformation is not bijective since it is not one-to-one.

Bijective transformations play a particularly important role.

Definition 1.3.8. *Let A be a set. A bijective transformation of A is called a permutation of A. The set of all permutations of A is denoted by* $\mathbf{S}(A)$. *Thus* $\phi \in \mathbf{S}(A)$ *if and only if* $\phi : A \longrightarrow A$ *is a bijective mapping.*

The word "permutation" has an alternative meaning since it is also widely used in combinatorics giving us a situation when the same word represents two different things, possibly leading to ambiguity. However these two ideas are closely connected and usually it is clear from the context which meaning of the term permutation is being used.

Let A be a finite set, say $A = \{a_1, a_2, \ldots, a_n\}$. Here, the order of the elements is not important. If π is a permutation of the set A, it can be represented in the following way

$$\begin{array}{cccccc} a_1 & a_2 & a_3 & \ldots & a_{n-1} & a_n \\ \downarrow & \downarrow & \downarrow & \ldots & \downarrow & \downarrow \\ \pi(a_1) & \pi(a_2) & \pi(a_3) & \ldots & \pi(a_{n-1}) & \pi(a_n) \end{array}$$

This can be considered as a renumeration of the elements of A.

Example. Let $A = \{1,2,3,4\}$ and consider the permutation π given by the chart below

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 4 & 3 & 2 & 1 \end{array}.$$

Now consider the set $\bar{A} = \{a_1, a_2, a_3, a_4\}$ and define the permutation $\bar{\pi}$ on $\bar{A}$ by the chart

$$\begin{array}{cccc} a_1 & a_2 & a_3 & a_4 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ a_4 & a_3 & a_2 & a_1 \end{array}.$$

Evidently, the first chart π could be used to represent the transformation $\bar{\pi}$ of the set $\bar{A}$ so we can represent a transformation of a set by indexing its elements and then tracking the changes in this indexing generated by the transformation.

In the same way, a permutation of any indexed set can be represented by the corresponding change in the indices. Since any finite set can be indexed, this gives us an easy way to represent such transformations. In this case the order of the elements in the set is important and is defined by the indexing. We shall denote a permutation of the finite set $\bar{A} = \{a_1, a_2, \ldots, a_n\}$ by an ordered tuple consisting of all the elements of $\bar{A}$ once and only once. We also say that this is a permutation of the elements $a_1, a_2, \ldots, a_n$. The elements in a tuple appear in some order: the tuple has a first element (unless it is empty), a second element (unless its length is less than 2), and so on. For example, if $\bar{A} = \{1,2,3\}$, then $(1,2,3)$ and $(3,2,1)$ are two different ways to list the elements of A in some order. These give two permutations of the numbers $1,2,3$.

Let $A = \{a_1, a_2, \ldots, a_n\}$ be a finite set with n elements, and let π be a permutation of the set A. Based on the considerations above, permutations of the set $\bar{A} = \{1,2,\ldots,n\}$ can be considered instead of permutations of the abstract set $A = \{a_1, a_2, \ldots, a_n\}$. Earlier we used the notation $\mathbf{S}(A)$ for the set of permutations of A. However, we will use the notation $\mathbf{S}_n$, or $\mathbf{Sym}(n)$, for the set of all permutations of the set $\{1,2,\ldots,n\}$. If $\pi \in \mathbf{S}_n$, then we will say that π is *a permutation of degree n*. The number of different permutations of the elements of the set A consisting of n elements is easily seen to be equal to $n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n$. Hence $|\mathbf{S}_n| = n!$

The permutation $\pi : \{1,2,\ldots,n\} \longrightarrow \{1,2,\ldots,n\}$ can be written as

$$\begin{pmatrix} 1 & 2 & \ldots & n \\ \pi(1) & \pi(2) & \ldots & \pi(n) \end{pmatrix}$$

which we will call the *tabular form of the permutation*. Since π is a permutation of the set $\{1,2,\ldots,n\}$ we see that

$$\{1,2,\ldots,n\} = \{\pi(1),\pi(2),\ldots,\pi(n)\}.$$

Thus the second row of a tabular form is a permutation of the numbers $1,2,\ldots,n$. It is not necessary to write all elements of the first row in the natural order from 1 to n, although this is usually the way such permutations are written. Sometimes it is convenient to write the first row in a different order. What is most important is that every element of the second row is the image of the corresponding element of the first row situated just above.

For example,

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 9 & 1 & 7 & 8 & 3 & 5 & 2 & 6 \end{pmatrix} \text{ and } \begin{pmatrix} 2 & 5 & 7 & 1 & 9 & 3 & 6 & 4 & 8 \\ 9 & 8 & 5 & 4 & 6 & 1 & 3 & 7 & 2 \end{pmatrix}$$

are the same permutation. Perhaps, for beginners, in order to better understand permutations, it may be worthwhile to write the permuation with arrows connecting the element of the first row with its image in the second row as in $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 4 & 9 & 1 & 7 & 8 & 3 & 5 & 2 & 6 \end{pmatrix}$. This method of writing a permutation should be quickly learned and then the student should revert to the shorthand notation.

We will multiply permutations π and σ by using the general rule of multiplication of mappings, namely composition of functions, introduced earlier in this section, with one important modification, suggested earlier. We note that normally we write and read from left to right. Thus, in writing the product $\pi \circ \sigma$ we first write π and then σ. Thus it is entirely natural that the first permutation to act should be the one that is written first, and after that the permutation that acts second is written and so on. Thus **for permutations only** when we write $\pi \circ \sigma$ we will mean that first the permutation π is performed and then the permutation σ. We remark that this is a personal preference and that in some books $\pi \circ \sigma$ means that first σ is performed and then π. This slight inconsistency is the result of writing mappings on the left; some algebra books write mappings on the right to avoid this.

According to this rule, the product of the two permutations π and σ is the permutation $\pi \circ \sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ \sigma(\pi(1)) & \sigma(\pi(2)) & \ldots & \sigma(\pi(n)) \end{pmatrix}$.

To multiply the two permutations in tabular form, in the first row of the table corresponding to the permutation π we choose an arbitrary element i. We obtain $\pi(i)$ from the second row of π corresponding to i. Then we find

this number $\pi(i)$ in the first row of the table corresponding to the permutation σ. In the second row of the second table just under this number $\pi(i)$ we find the number $\sigma(\pi(i))$. This is the image of i under the product $\pi \circ \sigma$. We can write it using the following convenient scheme:

$$\begin{array}{cccc} 1 & 2 & \dots & n \\ \downarrow & \downarrow & \dots & \downarrow \\ \pi(1) & \pi(2) & \dots & \pi(n) \\ \downarrow & \downarrow & \dots & \downarrow \\ \sigma(\pi(1)) & \sigma(\pi(2)) & \dots & \sigma(\pi(n)) \end{array} .$$

To illustrate this we give the following example, where the permutation π is written and done first.

Example. Let $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 1 & 5 & 2 \end{pmatrix}$ and $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \end{pmatrix}$. Then

$$\pi \circ \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 1 & 5 & 2 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 2 & 1 & 3 \end{pmatrix},$$

but

$$\sigma \circ \pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 1 & 5 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 2 & 4 \end{pmatrix}.$$

It is clearly the case that $\pi \circ \sigma \neq \sigma \circ \pi$. Hence multiplication of permutations is not a commutative operation.

The identity permutation is written as $\begin{pmatrix} 1 & 2 & 3 & \dots & n \\ 1 & 2 & 3 & \dots & n \end{pmatrix}$. Since a permutation is a bijection, each permutation has an inverse. It is easy to obtain the inverse permutation of a given permutation. All that we need for that is to interchange the upper row with the lower one, and then list the entries in the upper row in ascending order, making the corresponding position change in the bottom row of elements.

Example. Find the inverse of the permutation

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \end{pmatrix}.$$

First flip the upper and lower rows and then rearrange the elements of the upper row in ascending order:

$$\pi^{-1} = \begin{pmatrix} 2 & 3 & 4 & 5 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 2 & 3 & 4 \end{pmatrix}.$$

It is easy to check that the permutation obtained is the inverse element for π as we see below since

$$\begin{pmatrix}1&2&3&4&5\\2&3&4&5&1\end{pmatrix}\circ\begin{pmatrix}1&2&3&4&5\\5&1&2&3&4\end{pmatrix}=\begin{pmatrix}1&2&3&4&5\\1&2&3&4&5\end{pmatrix},$$

and

$$\begin{pmatrix}1&2&3&4&5\\5&1&2&3&4\end{pmatrix}\circ\begin{pmatrix}1&2&3&4&5\\2&3&4&5&1\end{pmatrix}=\begin{pmatrix}1&2&3&4&5\\1&2&3&4&5\end{pmatrix}.$$

Definition 1.3.9. *The permutation $\iota=\iota_{kt}$ of the set A is called a transposition (more precisely, the transposition of the symbols $k,t\in A$) if $\iota(k)=t,\iota(t)=k$, and $\iota(j)=j$ for all other elements $j\in A$.*

Thus a transposition interchanges two symbols and fixes the rest of them.

Example. $\begin{pmatrix}1&\mathbf{2}&3&\mathbf{4}&5\\1&\mathbf{4}&3&\mathbf{2}&5\end{pmatrix}$ is a transposition.

We say that the natural numbers *m,j form an inversion pair relative to the permutation* π, if $m<j$ but $\pi(m)>\pi(j)$. For **example**, the permutation $\begin{pmatrix}1&2&3&4\\1&4&3&2\end{pmatrix}$ contains three inversion pairs namely $(2,3),(2,4)$, and $(3,4)$

We let $\mathbf{inv}(\pi)$ denote the number of inversion pairs, relative to the permutation π. We define $\mathbf{sign}\,\pi=(-1)^{\mathbf{inv}(\pi)}$ and call $\mathbf{sign}\,\pi$ *the signature of the permutation* π.

In our last example, $\mathbf{sign}\,\pi=(-1)^3=-1$.

Definition 1.3.10. *The permutation π is called even, if $\mathbf{sign}\,\pi=1$ and π is called odd, if $\mathbf{sign}\,\pi=-1$. Thus π is even precisely when the number of inversion pairs of π is even and odd when the number of inversion pairs is odd.*

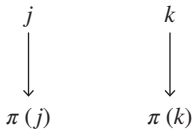
A short computation shows that the equation

$$\mathbf{sign}\,(\pi\circ\sigma)=\mathbf{sign}\,\pi\;\mathbf{sign}\,\sigma$$

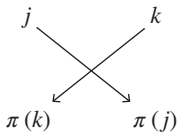
is valid for any permutations π and σ of the same degree. The equation $\mathbf{sign}\,(\pi\circ\sigma)=\mathbf{sign}\,\pi\;\mathbf{sign}\,\sigma$ implies that the product of two even permutations is even, the product of two odd permutations is even, and the product of an even and an odd permutation is odd.

There is a very convenient pictorial method for deciding whether a given permutation π is odd or even, based on the following observation. We rewrite

the permutation π as two rows of numbers, both in the order $1,2,\ldots,n$ and then draw a line from each number k to its image $\pi(k)$ in the second row. Let $1 \leq j < k \leq n$. If (j,k) is not an inversion pair then the two lines drawn from j to $\pi(j)$ and from k to $\pi(k)$ will not intersect. If the lines do intersect then this tells us that (j,k) is an inversion pair and the number of such crossovers for all pairs (j,k) determines the number of these. If numbers j and k don't form an inversion pair relative to π, we obtain a picture of the following type:

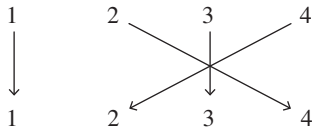


with no crossover of the corresponding lines. If numbers j and k make an inversion pair relative to π, we will have the following picture:



The total number of intersections of these lines is the number of inversion pairs.

We will illustrate this with the following example. Use the permutation we already used above: $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix}$. The following diagram corresponds to this permutation:



As we can see, there are three intersections corresponding to the three pairs of indices forming inversions $(2,3)$, $(2,4)$, and $(3,4)$. In practice we often write the permutation π in the usual fashion, the first row consisting of the elements $\{1,2,\ldots,n\}$ listed in that order. Then we draw lines from each number in the upper row *to the same number* in the bottom row. This is clearly equivalent to the procedure described above.

We let $\mathbf{A}_n$ denote the subset of $\mathbf{S}_n$ consisting of all even permutations. It is not difficult to prove that $|\mathbf{A}_n| = \frac{n!}{2}$.

The representation of permutations as products of cycles plays an important role in their study and we briefly discuss this idea next.

Definition 1.3.11. *Let $1 \leq r \leq n$. A permutation π is called a cycle of length k if there are natural numbers $j_1, \ldots, j_k$ such that*

$$\pi(j_1) = j_2, \pi(j_2) = j_3, \ldots, \pi(j_{k-1}) = j_k, \pi(j_k) = j_1.$$

and $\pi(s) = s$ for all $s \notin \{j_1, \ldots, j_k\}$. The cycle is denoted by $(j_1 j_2 \ldots j_k)$. The numbers $j_1, \ldots, j_k$ are called the elements of this cycle.

In other words, the permutation π "cycles" the indices $j_1, j_2, \ldots, j_k$ around (thus $j_1 \mapsto j_2 \mapsto j_3 \mapsto \ldots \mapsto j_r \mapsto j_1$) but leaves all other indices fixed.

For **example**,

$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 3 & 7 & 5 & 6 & 4 \end{pmatrix}$ is the cycle (47).

The identity permutation is written as the cycle $(1) = (2) = \ldots$ of length 1. The cycles of length 2 are precisely the transpositions. Notice also that, in this notation, it does not matter which j_r is listed first. Thus permuting the elements of a cycle in cyclic order gives us the same permutation, as for example $(235) = (352) = (523)$.

It is easy to check that cycles with no elements in common (e.g., (13) and (245)) commute with each other and therefore the order of writing the factors is not important in such a case. The following theorem illustrates the importance of cycles.

Theorem 1.3.12. *Every permutation can be represented as a product of cycles with no elements in common and this representation is unique to within the order of the factors.*

We shall not prove this theorem but illustrate the idea of the proof using the following example. Let

$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 3 & 2 & 1 \end{pmatrix}$. We see that π transforms 1 to 4, 4 to 3, 3 to 6, 6 to 1, which means that the cycle (1436) is part of the product decomposition. The permutation π transforms the remaining number 2 to 5, and 5 to 2. Therefore the transposition (25) is also part of the decomposition. So $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 3 & 2 & 1 \end{pmatrix} = (1436)(25) = (25)(1436)$. The reader can now probably imagine how a general proof would work.

Exercise Set 1.3

In each of the following questions explain your reasoning, either by giving a proof of your assertion or a counterexample.

1.3.1. Let A be a nonempty set. Prove that A is infinite if and only if $\mathbf{S}(A)$ is infinite.

1.3.2. Prove that there is a bijective mapping from $A \times B$ to $B \times A$.

1.3.3. Let A be a set consisting of two elements. Is the multiplication on the set $\mathbf{S}(A)$ commutative?

1.3.4. Let f: $\mathbb{N} \longrightarrow \mathbb{Z}$ be a mapping defined by the rule

$$f(n) = \begin{cases} \dfrac{n}{2} - 1 \text{ whenever } n \text{ is even,} \\ -\dfrac{n+1}{2} \text{ whenever } n \text{ is odd.} \end{cases}$$

Is f injective? If yes, find an inverse to f.

1.3.5. Let $f : \mathbb{Q} \longrightarrow \mathbb{Q}$ be a mapping defined by the rule $f(x) = 3x - |x|$, for $x \in \mathbb{Q}$. Is f injective? If yes, find an inverse for f.

1.3.6. Let $f : \mathbb{Q} \longrightarrow \mathbb{Q}$ be the mapping defined by $f(x) = 2x + |x|$, for $x \in \mathbb{Q}$. Is f injective? If yes, find an inverse for f.

1.3.7. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the mapping defined by

$$f(x) = \begin{cases} x^2 \text{ whenever } x \geq 0, \\ x(x-3) \text{ whenever } x < 0. \end{cases}$$

Is f injective? If yes, find an inverse for f.

1.3.8. Let $f : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N}$ be the mapping defined by $f(n,m) = 2^{n-1}(2m-1)$. Is f injective? If yes, find an inverse for f.

1.3.9. Let $f : \mathbb{Q} \longrightarrow \mathbb{Q}$ be the mapping defined by $f(x) = x^2 + 2$, and let $g : \mathbb{Q} \longrightarrow \mathbb{Q}$ be a mapping defined by $g(x) = \frac{x}{2} - 2$. Find the products $g \circ f, f \circ g, (f \circ g) \circ f$, and $f \circ (g \circ f)$

1.3.10. Let $f : \mathbb{Q} \longrightarrow \mathbb{Q}$ be the mapping defined by $f(x) = (1 + (1-x)^{\frac{1}{3}})^{\frac{1}{5}}$. Represent f as a product of four mappings.

1.3.11. Let $f : A \longrightarrow B, g : A \longrightarrow C$. Prove that if f, g are injective then so is $f \circ g$ and that if f, g are surjective then so is $f \circ g$.

1.3.12. Two cycles $(a_1\, a_2\, a_3 \ldots a_n)$ and $(b_1\, b_2\, b_3 \ldots b_r)$ are disjoint if $\{a_1, a_2, a_3, \ldots, a_n\} \cap \{b_1, b_2, b_3, \ldots, b_r\} = \emptyset$. Prove that disjoint cycles commute.

1.3.13. Write $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 3 & 6 & 5 & 11 & 7 & 9 & 8 & 1 & 10 & 2 & 4 \end{pmatrix}$ as a product of disjoint cycles and then as a product of transpositions.

1.3.14. Represent the permutation $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 6 & 4 & 5 & 7 & 2 & 8 & 3 & 9 & 1 \end{pmatrix}$ as a product of transpositions and find whether it is even or odd.

1.3.15. Find whether the permutation $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 3 & 1 & 4 & 7 & 5 & 6 \end{pmatrix}$ is even or odd.

1.3.16. Write $(123)(45)(1543)(276)$ first as a product of disjoint cycles, then as a product of transpositions, and then find whether it is even or odd.

1.3.17. Let α be a cycle of length r. Prove that $\alpha^r = \varepsilon$ and that r is the least natural number for which this is true.

1.3.18. If α and β are disjoint cycles of lengths r, s, respectively, then prove that $(\alpha\beta)^l = \varepsilon$, where l is the least common multiple of r and s. Prove also that l is the least natural number for which this is true.

1.3.19. Find the inverse of the permutation $(a_1 a_2 \ldots a_k)$.

1.3.20. Find $\alpha \circ \beta$ if

$$\alpha = (13)(1468)(26754) \text{ and } \beta = (356)(275)(8941).$$

1.4 OPERATIONS ON MATRICES

In this section we construct some useful examples—matrices—which can be used to illustrate the most important concepts of abstract algebraic structures. Additionally, however, matrices are one of the most useful and prevalent objects in mathematics and its applications. The language of matrices is very convenient and efficient, so is used by scientists everywhere. Matrices are also a central concept in linear algebra, which itself is useful in many fields.

An $m \times n$ matrix is a rectangular table of entries (or elements), containing m rows and n columns which may be numbers or, more generally, any abstract quantities that can be added and multiplied. If the number of rows is equal to the number of columns, then the matrix is called a *square (or quadratic) matrix*, and the number n of its rows (or columns) is called *the order of the matrix*. A matrix of order n is also called an $n \times n$ matrix, the first n refers to the number of rows and the second one to the number of columns. In this book we will mostly consider square matrices. Usually the matrices we consider will have at least order 2.

The choice of the entries used in a matrix depends on the branch of science in which they are used and on the specific problems to be solved. They could be numbers, or polynomials, or functions, or elements of some abstract algebraic structure. In this book we mostly consider numerical matrices, those matrices with numbers as elements.

We denote the entries of a matrix using lower case letters with two indices, which can be thought of as the coordinates of an element in the matrix. The first index shows the number of the row in which the element is situated, while the second index is the number of the place of the element in this row, or, which is the same, the number of the column in which the element lies. Thus an $n \times n$ matrix has the following form:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2,n-1} & a_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{n,n-1} & a_{nn} \end{pmatrix};$$

square brackets may also be used as in

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2,n-1} & a_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{n,n-1} & a_{nn} \end{bmatrix}.$$

We call a_{ij} the (i,j) entry of the matrix, so we list the row index first and the column index second. Thus a_{ij} is the entry of the matrix in row i, column j. We also shall use the following brief form for matrix notation

$$[a_{ij}]_{1 \le i,j \le n} \text{ or } [a_{ij}],$$

when the order is reasonably clear.

The set of $n \times n$ matrices whose entries belong to some set S will be denoted by $\mathbf{M}_n(S)$. In this book, we think of S as being a subset of the set, $\mathbb{R}$, of real numbers. In this case, we shall sometimes say that we are dealing with numerical matrices.

We make the following definition of equality of matrices.

Definition 1.4.1. *Two matrices*

$$A = [a_{ij}] \text{ and } B = [b_{ij}]$$

in the set $\mathbf{M}_n(S)$ *are said to be equal, if* $a_{ij} = b_{ij}$ *for every pair of indices* (i,j), *where* $1 \le i,j \le n$.

Thus, equal matrices should have the same order and the same elements in the corresponding places. Certain special types of matrices occur frequently and we next define some of these.

Definition 1.4.2. *Let $A = [a_{ij}]$ be an $n \times n$ numerical matrix.*

(i) *A is called upper triangular, if $a_{ij} = 0$ whenever $i > j$;*
(ii) *If A is upper triangular then A is called unitriangular, if $a_{ii} = 1$ for each $i, 1 \leq i \leq n$;*
(iii) *If A is upper triangular then A is called zero-triangular if $a_{ii} = 0$ for each $i, 1 \leq i \leq n$;*
(iv) *A is called diagonal, if $a_{ij} = 0$ for every $i \neq j$.*

For **example**, the matrix $\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix}$ is *upper triangular*, the matrix $\begin{pmatrix} 1 & a_{12} & a_{13} \\ 0 & 1 & a_{23} \\ 0 & 0 & 1 \end{pmatrix}$ is *unitriangular*, the matrix $\begin{pmatrix} 0 & a_{12} & a_{13} \\ 0 & 0 & a_{23} \\ 0 & 0 & 0 \end{pmatrix}$ is *zero-triangular*, and the matrix $\begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}$ is *diagonal*.

The power of matrices is perhaps best utilized as a means of storing information. An important part of this is concerned with certain natural operations defined on matrices, which we consider next. Just as we can build an arithmetic of numbers so we can build an arithmetic of matrices.

Definition 1.4.3. *Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be matrices in the set $\mathbf{M}_n(\mathbb{R})$. The sum $A + B$ of these matrices is the matrix $C = [c_{ij}] \in \mathbf{M}_n(\mathbb{R})$, whose entries are $c_{ij} = a_{ij} + b_{ij}$ for every pair of indices (i,j), where $1 \leq i,j \leq n$.*

Here is a very easy example to illustrate matrix addition.

Example. $\begin{pmatrix} 1 & 3 \\ 5 & 2 \end{pmatrix} + \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix} = \begin{pmatrix} 1+2 & 3+1 \\ 5+4 & 2+3 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 9 & 5 \end{pmatrix}$.

The definition means that we can only add matrices if they have the same order and then to add two matrices of the same order we just add the corresponding entries of the two matrices. In this way matrix addition is reduced to the addition of the corresponding entries. Therefore the operation of matrix addition inherits all the properties of number addition.

For example, let $A, B, C \in \mathbf{M}_n(\mathbb{R})$. Then addition of matrices is commutative, which means that $A+B = B+A$. This follows since $a_{ij}+b_{ij} = b_{ij}+a_{ij}$, where $A = [a_{ij}], B = [b_{ij}]$. Likewise, addition of matrices is associative, which means that $(A+B)+C = A+(B+C)$. The set $\mathbf{M}_n(\mathbb{R})$ has a zero matrix O each of whose entries is 0. The matrix O is called the (additive) identity element since $A+O = A = O+A$ for each matrix $A \in \mathbf{M}_n(\mathbb{R})$. It is not hard to see that for each n there is precisely one $n \times n$ matrix with the property that when it is added to A the result is again A. If $A = [a_{ij}]$ then the $n \times n$ matrix $-A$ is the matrix whose entries are $-a_{ij}$. It is easy to see from the definition of matrix addition that $A+(-A) = O = -A+A$. This matrix $-A$ is the unique matrix with the property that when it is added to A the result is the matrix O. The matrix $-A$ is called the additive inverse of the matrix A. Matrix subtraction can be introduced in $\mathbf{M}_n(\mathbb{R})$ by using the natural rule that $A-B = A+(-B)$. This amounts to simply subtracting corresponding entries of the matrix.

Compared to addition, matrix multiplication looks more sophisticated, and does not seem as natural as addition.

Definition 1.4.4. *Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two matrices in the set $\mathbf{M}_n(\mathbb{R})$. The product AB of these matrices is the matrix $C = [c_{ij}]$, whose elements are*

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}$$

for every pair of indices (i,j), where $1 \le i,j \le n$.

Thus to obtain the (i,j) entry of the product C, we need to multiply pairwise the elements of row i of the matrix A by the corresponding elements of column j of the matrix B and add the results.

Example. Let $A = \begin{pmatrix} 1 & 3 \\ 5 & 2 \end{pmatrix}, B = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}$,

$$\begin{pmatrix} \mathbf{1} & \mathbf{3} \\ 5 & 2 \end{pmatrix}\begin{pmatrix} \mathbf{2} & 1 \\ \mathbf{4} & 3 \end{pmatrix} = \begin{pmatrix} 1\times 2+3\times 4 = 14 & \\ & \end{pmatrix},$$

$$\begin{pmatrix} 1 & 3 \\ \mathbf{5} & \mathbf{2} \end{pmatrix}\begin{pmatrix} \mathbf{2} & 1 \\ \mathbf{4} & 3 \end{pmatrix} = \begin{pmatrix} 1\times 2+3\times 4 = 14 & \\ 5\times 2+2\times 4 = 18 & \end{pmatrix},$$

$$\begin{pmatrix} \mathbf{1} & \mathbf{3} \\ 5 & 2 \end{pmatrix}\begin{pmatrix} 2 & \mathbf{1} \\ 4 & \mathbf{3} \end{pmatrix} = \begin{pmatrix} 1\times 2+3\times 4 = 14 & 1\times 1+3\times 3 = 10 \\ 5\times 2+2\times 4 = 18 & \end{pmatrix},$$

$$\begin{pmatrix} 1 & 3 \\ \mathbf{5} & \mathbf{2} \end{pmatrix}\begin{pmatrix} 2 & \mathbf{1} \\ 4 & \mathbf{3} \end{pmatrix} = \begin{pmatrix} 1\times 2+3\times 4 = 14 & 1\times 1+3\times 3 = 10 \\ 5\times 2+2\times 4 = 18 & 5\times 1+2\times 3 = 11 \end{pmatrix}.$$

So, $\begin{pmatrix}1 & 3\\ 5 & 2\end{pmatrix}\begin{pmatrix}2 & 1\\ 4 & 3\end{pmatrix} = \begin{pmatrix}14 & 10\\ 18 & 11\end{pmatrix}$.

Matrix multiplication is not commutative as the previous example shows since

$$\begin{pmatrix}2 & 1\\ 4 & 3\end{pmatrix}\begin{pmatrix}1 & 3\\ 5 & 2\end{pmatrix} = \begin{pmatrix}7 & 8\\ 19 & 18\end{pmatrix} \neq \begin{pmatrix}1 & 3\\ 5 & 2\end{pmatrix}\begin{pmatrix}2 & 1\\ 4 & 3\end{pmatrix}.$$

The matrices A, B satisfying $AB = BA$ are called permutable (in other words these matrices commute), and not all matrices have this property. However, matrix multiplication does possess other important properties, namely, the associative and distributive properties, which are exhibited in the following theorem.

Theorem 1.4.5. *For arbitrary matrices $A, B, C \in \mathbf{M}_n(\mathbb{R})$, the following properties hold:*

(i) $(AB)C = A(BC)$;
(ii) $(A+B)C = AC + BC$;
(iii) $A(B+C) = AB + AC$;
(iv) *There exists a matrix $I = I_n \in \mathbf{M}_n(\mathbb{R})$ such that $AI = IA = A$ for each matrix $A \in \mathbf{M}_n(\mathbb{R})$. For a given value of n, I is the unique matrix with this property.*

Proof.

(i) By definition, we need to show that the corresponding entries of $(AB)C$ and $A(BC)$ are equal. To do this, let

$$A = [a_{ij}], B = [b_{ij}], \text{ and } C = [c_{ij}].$$

Put

$$AB = [d_{ij}], BC = [v_{ij}],$$
$$(AB)C = [u_{ij}], A(BC) = [w_{ij}].$$

We must show that $u_{ij} = w_{ij}$ for arbitrary (i, j), where $1 \leq i, j \leq n$. We have

$$u_{ij} = \sum_{1\leq k\leq n} d_{ik}c_{kj} = \sum_{1\leq k\leq n}\left(\sum_{1\leq m\leq n} a_{im}b_{mk}\right)c_{kj} = \sum_{1\leq k\leq n}\sum_{1\leq m\leq n}(a_{im}b_{mk})c_{kj}$$

and

$$w_{ij} = \sum_{1\le m\le n} a_{im}v_{mj} = \sum_{1\le m\le n} a_{im}\left(\sum_{1\le k\le n} b_{mk}c_{kj}\right)$$
$$= \sum_{1\le m\le n}\sum_{1\le k\le n} a_{im}(b_{mk}c_{kj}) = \sum_{1\le k\le n}\sum_{1\le m\le n} a_{im}(b_{mk}c_{kj}).$$

Since $(a_{im}b_{mk})c_{kj} = a_{im}(b_{mk}c_{kj})$ it follows that $u_{ij} = w_{ij}$ for all pairs (i,j). Hence $(AB)C = A(BC)$.

(ii) We need to show that corresponding entries of $AC+BC$ and $A(B+C)$ are equal. Put

$$AC = [x_{ij}], BC = [y_{ij}], (A+B)C = [z_{ij}].$$

We shall prove that $z_{ij} = x_{ij}+y_{ij}$ for arbitrary i,j, where $1 \le i,j \le n$. We have

$$z_{ij} = \sum_{1\le k\le n} (a_{ik}+b_{ik})c_{kj} = \sum_{1\le k\le n} a_{ik}c_{kj} + \sum_{1\le k\le n} b_{ik}c_{kj} = x_{ij}+y_{ij}.$$

Thus $(A+B)C = AC+BC$.

The proof of (iii) is similar.

(iv) We define the symbol δ_{ij} (the Kronecker delta) by

$$\delta_{ij} = \begin{cases} 0, \text{ if } i \ne j, \\ 1, \text{ if } i = j. \end{cases}$$

It is easy to check that

$$I = [\delta_{ij}] = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

has the required property that $AI = IA = A$.

In order to prove the uniqueness of I assume that there also exists a matrix U such that $AU = UA = A$ for each matrix $A \in \mathbf{M}_n(\mathbb{R})$. Setting $A = I$ we obtain $IU = I$. Also, though, we know that $IU = U$, from the definition of I, so that $I = U$.

The matrix $I = I_n$ is called the $n \times n$ *identity matrix*.

Definition 1.4.6. *Let $A \in \mathbf{M}_n(\mathbb{R})$. The matrix $B \in \mathbf{M}_n(\mathbb{R})$ is called an inverse (or reciprocal) of A if $AB = BA = I$. The matrix A is then said to be invertible or non-singular.*

Many nonzero matrices lack inverses. For example, consider the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. Then for an arbitrary matrix $\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$ we have $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} \\ 0 & 0 \end{pmatrix}$. This product would never be the identity matrix, thus our matrix has no inverse.

If a matrix A has an inverse, then this inverse matrix is unique. To see this let U, V be inverses of the matrix A so that $AU = UA = I = VA = AV$. Then we have

$$V(AU) = VI = V \text{ and } V(AU) = (VA)U = IU = U.$$

Thus $V = U$. We follow the usual convention and denote the inverse of the matrix A by A^{-1}.

We note that criteria for the existence of an inverse for a given matrix are closely connected to the idea of the determinant of a matrix, a concept usually introduced in a linear algebra course, where the properties of determinants are investigated. We shall need only one result, which states that a matrix $A \in \mathbf{M}_n(\mathbb{R})$ has a multiplicative inverse if and only if its determinant, $\mathbf{det}(A)$, is nonzero. The matrix A is called nonsingular in this case and singular if $\mathbf{det}(A) = 0$. The proof can be found in any treatise on linear algebra.

Now we consider multiplication of a matrix by a number, or scalar.

Definition 1.4.7. *Let $A = [a_{ij}]$ be a matrix from the set $\mathbf{M}_n(\mathbb{R})$ and let $\alpha \in \mathbb{R}$. The product of α and the matrix A is the matrix $\alpha A = [c_{ij}] \in \mathbf{M}_n(\mathbb{R})$, whose entries are defined by $c_{ij} = \alpha a_{ij}$, for every pair of indices (i,j), where $1 \le i, j \le n$.*

Thus, when we multiply a matrix by a real number we multiply each element of the matrix by this number. Here are the main properties of this operation, which can be proved quite easily, in a manner similar to that given in Theorem 1.4.5. We note that these equations hold for all real numbers α, β and for all matrices A, B where the multiplication is defined.

Theorem 1.4.8. *Let A, B be matrices and α, β real numbers.*

(i) $(\alpha + \beta)A = \alpha A + \beta A$;
(ii) $\alpha(A + B) = \alpha A + \alpha B$;
(iii) $\alpha(\beta A) = (\alpha\beta)A$;

(iv) $1A = A$;
(v) $\alpha(AB) = (\alpha A)B = A(\alpha B)$.

Note that this operation of multiplying a matrix by a number can be reduced to the multiplication of two matrices since $\alpha A = (\alpha I)A$.

Here is a summary of all the properties we have obtained so far, using our previously established notation.

$A+B = B+A$,
$A+(B+C) = (A+B)+C$,
$A+O = A$,
$A+(-A) = O$,
$A(B+C) = AB+AC$,
$(A+B)C = AC+BC$,
$A(BC) = (AB)C$,
$AI = IA = A$,
$(\alpha+\beta)A = \alpha A+\beta A$,
$\alpha(A+B) = \alpha A+\alpha B$,
$\alpha(\beta A) = (\alpha\beta)A$,
$1A = A$,
$\alpha(AB) = (\alpha A)B = A(\alpha B)$.

Exercise Set 1.4

In each of the following questions explain your reasoning, either by giving a proof of your assertion or a counterexample.

1.4.1. Let A be a diagonal matrix. Suppose all entries on the main diagonal are different. Let B be a matrix such that $AB = BA$. Prove that B is diagonal.

1.4.2. Find all matrices $A \in M_2(\mathbb{R})$ with the property $A^2 = O$.

1.4.3. Let A and B be matrices. If we interchange the m-th and t-th rows of A, what changes does this imply in the matrix AB?

1.4.4. Let $A, B \in \mathbf{M}_n(\mathbb{R})$. If $\alpha \in \mathbb{R}$ and if we add α times row t to row m in the matrix A then what changes does this imply in the matrix AB?

1.4.5. Find A^3 if $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

1.4.6. Find $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}^3$.

1.4.7. Find $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}^3$.

1.4.8. Find $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ 4 & 0 & 0 & 1 \end{pmatrix}^3$.

1.4.9. Find $\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}^4$.

1.4.10. Let $A \in \mathbf{M}_n(\mathbb{R})$. A matrix A is called nilpotent, if $A^k = O$ for some positive integer k. The minimal such number k is called the nilpotency class of A. Prove that every zero triangular matrix is nilpotent.

1.4.11. Let $A \in \mathbf{M}_2(\mathbb{R})$ be a matrix such that $AX = XA$ for all $X \in \mathbf{M}_2(\mathbb{R})$. Prove that $A = rI$ for some $r \in \mathbb{R}$. What will the general form of this result be?

1.4.12. If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and if $ad - bc \neq 0$ then show that $A^{-1} = \dfrac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

1.4.13. Solve the following matrix equation $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} X = \begin{pmatrix} 3 & 5 \\ 5 & 9 \end{pmatrix}$.

1.4.14. Find the matrix products: $\begin{pmatrix} 3 & -2 \\ 4 & -1 \end{pmatrix}\begin{pmatrix} -5 & 2 \\ 2 & 4 \end{pmatrix}$ and $\begin{pmatrix} 1 & -1 & 4 \\ 3 & -2 & 7 \\ -2 & 3 & -4 \end{pmatrix}\begin{pmatrix} -2 & 0 & 3 \\ 0 & -4 & 7 \\ 2 & -1 & 5 \end{pmatrix}$.

1.4.15. Prove that if A, B are invertible matrices of $\mathbf{M}_n(\mathbb{R})$ then AB is invertible and find a formula for its inverse in terms of A^{-1} and B^{-1}.

1.4.16. Prove that if $A \in \mathbf{M}_n(\mathbb{R})$ is such that $A^n = O$ then $I + A$ is invertible.

1.4.17. Find all matrices $A \in \mathbf{M}_2(\mathbb{R})$ such that $A^2 = I$.

1.4.18. If $A = [a_{ij}] \in \mathbf{M}_n(\mathbb{R})$ then the transpose of A is the matrix $A^t = [b_{ij}]$ where $b_{ij} = a_{ji}$. Show that $(AB)^t = B^t A^t$ whenever also $B \in \mathbf{M}_n(\mathbb{R})$.

1.4.19. A matrix $A = [a_{ij}] \in \mathbf{M}_n(\mathbb{R})$ is symmetric if $a_{ji} = a_{ij}$ for all $i \neq j$ and skew symmetric if $a_{ji} = -a_{ij}$ when $i \neq j$. Prove the following facts: (a) $A + A^t$ is symmetric, (b) $A - A^t$ is skew symmetric.

1.4.20. Prove that every square matrix is a sum of a symmetric matrix and a skew symmetric matrix, in a unique way.

1.5 BINARY ALGEBRAIC OPERATIONS AND EQUIVALENCE RELATIONS

In this section we are interested in binary (algebraic) operations; these are important in mathematics and certainly in modern algebra. Indeed, modern algebra could be regarded as a branch of mathematics that studies algebraic operations, since much of the time we are not interested in the nature of the elements of a set, but are more interested in how an algebraic operation defined on the set acts on the elements of the set.

The usual addition and multiplication of two rational numbers are just two examples of binary operations defined on the set $\mathbb{Q}$ of rational numbers. The main idea here is that associated with every ordered pair of rational numbers there is another rational, its sum or product. Thus addition and multiplication can really be thought of as mappings of the Cartesian product $\mathbb{Q} \times \mathbb{Q}$ to $\mathbb{Q}$.

As another example of this concept we recall that in Section 1.3 the set of permutations $\mathbf{S}_n$ was introduced for each natural number n. There we saw how to define the product of two permutations in $\mathbf{S}_n$, which we now view as a mapping from the Cartesian product $\mathbf{S}_n \times \mathbf{S}_n$ to $\mathbf{S}_n$ and this is also a binary operation on the set of such permutations. In this case, we recall that the multiplication is not commutative.

Definition 1.5.1. *Let M be a set. The mapping $\theta : M \times M \longrightarrow M$ from the Cartesian square of M to M is called a binary (algebraic) operation on the set M. Thus, corresponding to every ordered pair (a,b) of elements, where $a, b \in M$, there is a uniquely defined element $\theta(a,b) \in M$. The element $\theta(a,b) \in M$ is called the composition of the elements a and b relative to this operation.*

Notice that there are two important ideas here. One is that $\theta(a,b)$ is an element of M; the other is that $\theta(a,b)$ is uniquely determined by the ordered pair (a,b). Furthermore, it is often rather cumbersome to keep referring to the

function θ and using the notation $\theta(a,b)$. Therefore $\theta(a,b)$ is often written $a*b$ or as $a\circ b$ so that $*$ (or $\circ$, or some other symbol) denotes the operation. In many natural cases where addition is involved the operation will usually be denoted by + and the corresponding composition $a+b$ is then called the *sum* of a and b. In this case, we talk about the *additive notation* of the binary operation. Often also, multiplication is the operation involved and in this case the sign $\cdot$ is usually used for multiplication; the corresponding composition $a\cdot b$ is called the *product* of a and b. In this case, we say that *multiplicative notation* is being used. Traditionally, the $\cdot$ is omitted and we will also often follow this convention so that $a\cdot b$ will usually be writen as ab.

We now give some further examples of binary operations. Usually it is quite easy to verify that these are binary operations, but they serve to illustrate that binary operations are very familiar to the reader, as we observed above. The question to be resolved, for a given binary operation $*$ defined on a set M, is whether or not $a*b\in M$, for all $a,b\in M$.

(i) Addition on the sets $\mathbb{N},\mathbb{Z},\mathbb{Q},\mathbb{R}$;

(ii) Multiplication on the sets $\mathbb{N},\mathbb{Z},\mathbb{Q},\mathbb{R}$;

(iii) Let M be a set and let $\mathbf{P}(M)$ be the set of all transformations of M. Then, for all $f,g\in\mathbf{P}(M)$, the map θ defined by $\theta(f,g)=f\circ g$ is a binary operation on the set $\mathbf{P}(M)$;

(iv) Addition and multiplication of matrices in $\mathbf{M}_n(\mathbb{R})$.

(v) Addition and multiplication of real functions (i.e., transformations of the set $\mathbb{R}$).

(vi) Addition of vectors and vector product on the space $\mathbb{R}^3$.

(vii) The mappings$(n,k)\longmapsto n^k,(n,k)\longmapsto n^k+k^n,n,k\in\mathbb{N}$ define binary operations on $\mathbb{N}$.

(viii) The mappings $(n,k)\longmapsto\mathbf{GCD}(n,k)$, the greatest common divisor of n and k, and $(n,k)\longmapsto\mathbf{LCM}(n,k)$, the least common multiple of n and k, define binary operations on $\mathbb{Z}$. (Here the nonnegative $\mathbf{GCD}(n,k)$ and $\mathbf{LCM}(n,k)$ are chosen.)

We now consider some important properties of binary algebraic operations. To be concrete we may use the multiplicative form of writing a binary operation but may also illustrate the additive form. However we stress that our binary operations are very much more general than ordinary addition or multiplication.

Definition 1.5.2. *Let M be a set with a binary algebraic operation $*$. A subset S of M is called* closed *or* stable *with respect to $*$ if for each pair of elements $a,b\in S$ the element $a*b$ also belongs to S.*

This means that the restriction of the binary operation $*$ to S is a binary operation on S.

Example. The set $\mathbb{N}$ is a closed subset of $\mathbb{Z}$ with respect to addition, but not with respect to subtraction since, for example, $1,2$ are natural numbers but $1-2$ is not. The subset of all even integers is a closed subset of $\mathbb{Z}$ with respect to addition and multiplication, while the subset of odd integers is closed with respect to multiplication, but not addition.

Definition 1.5.3. *A binary operation on a set M is called commutative if $ab = ba$ for each pair a,b of elements of M.*

For the additive form, commutativity of a and b would be written as follows:

$$a+b = b+a, \text{ where } a,b \in M.$$

Concrete examples of commutative operations include: multiplication and addition on the set of integers, rational, and real numbers; matrix addition; multiplication of real functions; the operations **GCD**, **LCM** on $\mathbb{Z}$ and vector addition in $\mathbb{R}^3$. As we saw earlier, multiplication of transformations is not commutative in general. Likewise, multiplication of matrices is not commutative in general.

If we have three elements $a,b,c \in M$, then we can form the products $a(bc)$ and $(ab)c$ and in general these may be different as when we form $(a-b)-c$ and $a-(b-c)$, for real numbers a,b,c.

Definition 1.5.4. *A binary operation on a set M is called associative if $(ab)c = a(bc)$ for all elements a,b,c of M.*

Written additively this becomes

$$(a+b)+c = a+(b+c).$$

The examples mentioned, except for the examples involving the operations

$$(n,k) \longmapsto n^k, (n,k) \longmapsto n^k + k^n,$$

on $\mathbb{N}$, and the vector product on $\mathbb{R}^3$, are associative. Thus, when $n*k = n^k$ then $(2*1)*3 = 8$ whereas $2*(1*3) = 2$.

For four elements a,b,c,d, we can construct a number of different products. For example, we can determine each of the products

$$((ab)c)d, (ab)(cd), (a(bc))d, a(b(cd)) \text{ and } a((bc)d)$$

to name but a few. When the operation is associative, however, all methods of bracketing give the same expression so that there is no need for any complicated bracketing. For example, we have

$(a(bc))d = ((ab)c)d$;
$(ab)(cd) = ((ab)c)d$;
$a(b(cd)) = (ab)(cd) = ((ab)c)d$;
$a((bc)d) = (a(bc))d = ((ab)c)d$.

It can be shown in general that when a binary operation is associative the way in which we position the brackets in an expression makes no difference, assuming that the order of the elements is unchanged. In particular, we do not even need to put brackets in a product of elements $a_1, a_2, \ldots, a_n$ and just write the product as $a_1 a_2 \ldots a_n$ or, more succinctly, $\prod_{1 \le i \le n} a_i$. When needed we can place parentheses in any manner. In the case when $a_1 = a_2 = \cdots = a_n = a$, we will denote the product $a_1 a_2 \ldots a_n$ by a^n, as usual, and call it the n-th power of a.

For an associative binary operation on a set M the usual "rule of exponents" holds, at least for exponents that are natural numbers. Thus for each element $a \in M$ and arbitrary $n, m \in \mathbb{N}$ we have

$$a^n a^m = a^{n+m}, (a^n)^m = a^{nm}.$$

When additive notation is used, instead of multiplicative, powers become multiples; thus instead of $\prod_{1 \le i \le n} a_i$ we write $\sum_{1 \le i \le n} a_i$ and if $a_1 = a_2 = \cdots = a_n$, then write $a_1 + a_2 + \cdots + a_n = na$. In this case the rules of exponents become properties of multiples as follows:

$$na + ma = (n+m)a,\ m(na) = (mn)a.$$

Two elements a, b are said to *commute* or *permute* if $ab = ba$. We also sometimes say a and b are permutable. If the elements a, b commute then $(ab)^n = a^n b^n$ for each $n \in \mathbb{N}$. More generally, if $a_1, a_2, \ldots, a_n$ are elements of M and if the operation on M is commutative and associative, then

$$(a_1 a_2 \ldots a_n)^m = a_1^m a_2^m \ldots a_n^m$$

for every $m \in \mathbb{N}$. Additively this would be written as

$$m(a_1 + a_2 + \cdots + a_n) = ma_1 + ma_2 + \cdots + ma_n.$$

Definition 1.5.5. *Let M be a set with binary operation $*$. The element $e \in M$ is called a neutral (or identity) element under this operation if $a * e = e * a = a$ for each element a of the set M.*

The neutral element of a set M is unique whenever it exists. Indeed, if e_1 is another element with the property $a * e_1 = e_1 * a = a$ for all $a \in M$, then we may let $a = e$ or $a = e_1$, in the definitions of e_1 and e respectively, and then we obtain $e = e_1 * e = e_1$. Sometimes, to avoid ambiguity, we may use the notation e_M for the identity element of M.

If multiplicative notation is used then we use the term *identity element*, and often use the notation 1, or 1_M, for the neutral element e. In this case we have $1 \cdot a = a \cdot 1 = a$, for all $a \in M$. We emphasize that 1_M need not be the integer 1 here. If additive notation is used, then the neutral element is usually called the *zero element* and is often denoted by 0, or 0_M, so that the definition of the zero element is $a + 0 = 0 + a = a$ for each element $a \in M$; again 0_M should not be confused with the integer 0.

Examples.

(i) The operation of addition on the sets of all natural, integer, rational, and real numbers has a zero element, the number 0.

(ii) The operation of multiplication on the sets of all natural, integer, rational, and real numbers has an identity element,the number 1.

(iii) Let M be a set and $\mathbf{P}(M)$ be the set of all transformations of the set M. When the operation is composition of transformations of the set M, the identity element is the permutation $\varepsilon_M : M \longrightarrow M$, defined by $\varepsilon_M(m) = m$, for all $m \in M$.

(iv) The zero matrix is the zero element for the operation of addition on the set $\mathbf{M}_n(\mathbb{R})$ of real matrices whereas the identity matrix I is the identity element when the operation is multiplication on the set $\mathbf{M}_n(\mathbb{R})$.

(v) The function with value 0 for all elements in the domain is the zero element when real functions are added: and when the operation is multiplication the identity element is the function $f(x)$ for which $f(x) = 1$ for all $x \in \mathbb{R}$.

(vi) The operation $n * k = \mathbf{GCD}(n, k)$, whenever $n, k \in \mathbb{Z}$, has neutral element the number 0, since $\mathbf{GCD}(n, 0) = n$, for all $n \in \mathbb{Z}$.

(vii) For addition of vectors in $\mathbb{R}^3$, the zero element is the zero vector.

Definition 1.5.6. *Let M be a set with a binary operation and suppose that there is an identity element e. The element $x \in M$ is called an inverse of the element $a \in M$ if*

$$ax = xa = e.$$

If a has an inverse then we say that a is invertible.

When we use additive notation we will often also use the term "additive inverse" and when we use multiplicative notation we will also use the term "multiplicative inverse."

If the operation on M is associative and the element $a \in M$ is invertible then a has a unique inverse. To see this, let y be an element of M that also satisfies

$$ay = ya = e.$$

Then

$$y = ey = (xa)y = x(ay) = xe = x.$$

We denote the unique inverse of a by a^{-1}. We note that $aa^{-1} = a^{-1}a = e$ and so, evidently,

$$\left(a^{-1}\right)^{-1} = a.$$

If the operation on M is written additively then we denote the inverse of a, should it exist, by $-a$, called the negative (or sometimes the opposite) of a. In this case the definition of the additive inverse takes the form:

$$a + (-a) = -a + a = 0_M.$$

Proposition 1.5.7. *Let M be a set with an associative binary operation and suppose that M has an identity element e. If the elements $a_1, a_2, \ldots, a_n$ are invertible in M, then the product $a_1 a_2 \ldots a_n$ is also invertible and*

$$(a_1 a_2 \ldots a_n)^{-1} = a_n^{-1} a_{n-1}^{-1} \ldots a_1^{-1}.$$

Proof. We have, informally,

$$\begin{aligned}(a_1 a_2 \ldots a_n)(a_n^{-1} a_{n-1}^{-1} \ldots a_1^{-1}) &= (a_1 \ldots a_{n-1})(a_n a_n^{-1})(a_{n-1}^{-1} \ldots a_1^{-1}) \\ &= (a_1 a_2 \ldots a_{n-1})(a_{n-1}^{-1} \ldots a_1^{-1}) = \cdots = e.\end{aligned}$$

This shows that the proposition holds, since we have exhibited an element which multiplies $a_1 \ldots a_n$ to give e.

The existence of an identity element and the inverse of an element a allows us to define all integer powers of a. To do this we define

$$a^0 = e, \text{ and } a^{-n} = \left(a^{-1}\right)^n, \text{ whenever } n \in \mathbb{N}.$$

In additive notation these definitions take the form:

$$0a = 0_M \text{ and } (-n)a = n(-a).$$

Our next result shows that the usual rules of exponents hold for all integer powers.

Proposition 1.5.8. *Let M be a set together with an associative binary operation and suppose that M has an identity element e. If $a \in M$ is invertible and $m, n \in \mathbb{Z}$ then*

$$a^n a^m = a^{n+m} \text{ and } (a^n)^m = a^{nm}.$$

Proof. If $n, m > 0$, then the assertion follows by simply writing out the products. Furthermore if one of m or n is 0 then the equalities hold in any case. If $m, n < 0$, then $n = -p, m = -q$, for certain $p, q \in \mathbb{N}$. Then, using the definitions we have,

$$a^n a^m = a^{-p}a^{-q} = (a^{-1})^p (a^{-1})^q = (a^{-1})^{p+q} = a^{-(p+q)} = a^{-p-q} = a^{n+m}$$

and

$$(a^n)^m = (a^{-p})^{-q} = \left((a^p)^{-1}\right)^{-q} = \left(\left((a^p)^{-1}\right)^{-1}\right)^q = (a^p)^q = a^{pq} = a^{mn}.$$

Suppose now that $n > 0$, $-q = m < 0$ and $n > -m = q$. Then

$$a^n a^m = \underbrace{a \dots a}_{n}\underbrace{(a^{-1}) \dots (a^{-1})}_{q} = \underbrace{a \dots a}_{n-q} = a^{n+m}.$$

If $n > 0, -q = m < 0$ and $n < -m = q$, then

$$a^n a^m = \underbrace{a \dots a}_{n}\underbrace{(a^{-1}) \dots (a^{-1})}_{q} = \underbrace{a^{-1} \dots a}_{q-n}^{-1} = (a^{-1})^{-(n+m)} = a^{n+m}.$$

For the second equation, if $n > 0$ and $-q = m < 0$ then

$$(a^n)^m = ((a^n)^{-1})^q = ((a^{-1})^n)^q = (a^{-1})^{nq} = (a^{-1})^{-nm} = a^{-(-nm)} = a^{nm}.$$

If $-p = n < 0, m > 0$, then

$$(a^n)^m = ((a^{-1})^p)^m = (a^{-1})^{pm} = (a^{-1})^{-nm} = a^{-(-nm)} = a^{nm}.$$

The result follows.

Equivalence relations

In Section 1.2 we defined a binary relation on a set A to be a subset of the Cartesian product $A \times A$. If Φ is such a binary relation and if $(x, y) \in \Phi$, then we say that elements x and y (in this given order) correspond to each other via Φ. Instead of the notation $(x, y) \in \Phi$ we will use the notation $x \Phi y$, which is more suggestive, since typically we think of x and y being related by Φ. This form of notation is called *infix*.

Here are the most important properties of binary relations.

Definition 1.5.9. *Let A be a set with a binary relation Φ.*

(i) *Φ is called reflexive if $(a, a) \in \Phi$ (or $a \Phi a$), for each $a \in A$;*

(ii) *Φ is called transitive if, whenever $a, b, c \in A$ and $(a, b), (b, c) \in \Phi$, then $(a, c) \in \Phi$ (or, alternatively, $a \Phi b$ and $b \Phi c$ imply that $a \Phi c$);*

(iii) *Φ is called symmetric if, whenever $a, b \in A$ and $(a, b) \in \Phi$, then $(b, a) \in \Phi$ (or, alternatively, $a \Phi b$ implies $b \Phi a$);*

(iv) *Φ is called antisymmetric if, whenever $a, b \in A$ and $(a, b), (b, a) \in \Phi$ then $a = b$ (or, alternatively, $a \Phi b$ and $b \Phi a$ imply $a = b$).*

If A is a finite set we make a pair of perpendicular axes and label the axes with points representing the elements of A. If $a, b \in A$ and $a \Phi b$, then we can plot the point (a, b), as we do in the usual rectangular coordinate system, by finding the point on the horizontal axis labelled a and the point on the vertical axis labelled b and putting a mark (cross or circle) at the place where the lines drawn from these points would intersect. In this way, relations can be pictured.

There are many **examples** of reflexive relations and here we give only a few: If A is the set of all straight lines in the plane then the relation of "being parallel" is certainly reflexive; the relation "looks alike" on a certain set of people is clearly reflexive since everyone looks alike themselves; the relation of "having the same gender" on a set of animals is certainly reflexive and so on.

The relation "x is the brother of y" is symmetric on the set of all males, but is not symmetric on the set of all people, since y will only be the brother of x if y is a male. Here are some examples of transitive relations: the relation "to be divisible by" on the set of integers, the relation "to be greater" on the set of real numbers, the relation "to be older" on a set of people, the relation "to have the same color" on the set of toys, and so on.

The following concept leads us to an important type of binary relation called an equivalence relation.

Definition 1.5.10. *A family $\mathfrak{S}$ of subsets of a set A is called a covering if $A = \cup\mathfrak{S}$ (thus for each $x \in A$ there exists $S \in \mathfrak{S}$ such that $x \in S$). A covering $\mathfrak{S}$ is called a partition of the set A if, additionally, $X \cap Y = \emptyset$, whenever $X, Y \in \mathfrak{S}$ and $X \neq Y$; thus all pairs of distinct subsets of the partition have empty intersection.*

Let $\mathfrak{S}$ be a partition of the set A and define a binary relation $\Gamma(\mathfrak{S})$ on A by the rule that $(x,y) \in \Gamma(\mathfrak{S})$ if and only if the elements x and y belongs to the same set S from the family $\mathfrak{S}$. The relation $\Gamma(\mathfrak{S})$ has various properties which we now discuss. Since $A = \cup\mathfrak{S}$ then, for each element $x \in A$, there exists a subset $S \in \mathfrak{S}$ such that $x \in S$. Thus $(x,x) \in \Gamma(\mathfrak{S})$ and hence the relation $\Gamma(\mathfrak{S})$ is reflexive. It is clear that the relation $\Gamma(\mathfrak{S})$ is symmetric. Finally, let $(x,y), (y,z) \in \Gamma(\mathfrak{S})$. It follows that there exist subsets $S, R \in \mathfrak{S}$ such that $x, y \in S$ and $y, z \in R$. In particular, $y \in S \cap R$ and using the definition of a partition, we see that $S = R$. Hence the elements x, z belongs to S which is an element of the partition $\mathfrak{S}$. Thus $(x,z) \in \Gamma(\mathfrak{S})$ so the relation $\Gamma(\mathfrak{S})$ is transitive.

Definition 1.5.11. *A binary relation Φ on a set A is called an equivalence relation or an equivalence if it is reflexive, symmetric, and transitive.*

We give some examples next. First we say that two polygons are equivalent if they have the same number of vertices. Thus, for example, under this relation all triangles are equivalent, and it is easy to see that this relation is an equivalence relation. The family of all triangles can itself be partitioned into the subsets of acute, right-angled, and obtuse triangles and this partition helps define an equivalence relation on the set of all triangles. We can also say that two triangles are equivalent depending upon whether they are scalene, isosceles, or equilateral. Thus a given set may have more than one equivalence relation defined on it. More generally, the relation "the figure A is similar to the figure B" on the set of all geometric figures is an equivalence relation. We note too that our work here shows that every partition $\mathfrak{S}$ of a set A gives rise to an equivalence relation $\Gamma(\mathfrak{S})$ defined on A.

One main reason for studying equivalence relations is that such relations allow us to construct new mathematical objects quite rigorously. For example, the relation of colinearity of rays is a partition of the plane or space into classes of colinear rays. Each of these classes is called a direction, or a path. In this way we can transform the intuitive idea of direction into a rigorously defined concept. In a similar way, given a collection of figures we can define a relation on this set of figures by saying that figure A is related to figure B if and only if A has the same shape as B. Children forever use partitions (and hence equivalence relations!) in their play. For example a child might sort its toys according to color and the relation "is the same color as" is an equivalence relation.

Here is a list of some further **examples** of equivalence relations.

(i) If A is an arbitrary set there are two extreme cases: $\Phi = A \times A$ and $\Phi = \{(x,x) \mid x \in A\}$ (the diagonal of the Cartesian product $A \times A$). These are both examples of equivalence relations and all other equivalence relations on A are situated between these two extreme cases;
(ii) the relation "to be parallel" on the set of all straight lines in a plane;
(iii) the relation of similarity;
(iv) the relation "to be equivalent equations" on the set of equations;
(v) the relation "to belong to the same species" on the set of animals;
(vi) the relation "to be relatives" on the set of people;
(vii) the relation "to be the same height" on the set of people;
(viii) the relation "to live in the same city" on the set of people;
(ix) the relation "has the same birthday as" on the set of all people;
(x) the relation "is similar to" or "congruent to" on the set of all triangles;
(xi) the relation "has the same image" on the elements of the domain of a function.

We have already seen that each partition of a set generates an equivalence relation. We now show that, conversely, each equivalence relation on a set leads to a natural partition of the set.

Definition 1.5.12. *Let Φ be an equivalence relation on the set A and let $x \in A$. The subset $[x]_\Phi = \{y \in A \mid (x,y) \in \Phi\}$ is called the equivalence class of x.*

Thus the equivalence class of x consists precisely of those elements of A that are equivalent to x. It is important to note that each equivalence class is uniquely defined by each of its elements. Indeed, let $y \in [x]_\Phi$ so that $(x,y) \in \Phi$. If $z \in [y]_\Phi$, then $(y,z) \in \Phi$ also. Since the equivalence relation is transitive it follows that $(x,z) \in \Phi$ also and hence $z \in [x]_\Phi$. Thus $[y]_\Phi \subseteq [x]_\Phi$. Because equivalence relations are symmetric we also have $[x]_\Phi \subseteq [y]_\Phi$ and hence $[x]_\Phi = [y]_\Phi$.

Since $(x,x) \in \Phi$, it follows that $x \in [x]_\Phi$ and hence the family of all equivalence classes forms a covering set of A. Next we consider the intersection, $[x]_\Phi \cap [y]_\Phi$, of two equivalence classes and suppose that this intersection is not empty. Let $z \in [x]_\Phi \cap [y]_\Phi$. Then, as we noted above, $[z]_\Phi = [y]_\Phi$ and $[z]_\Phi = [x]_\Phi$ from which it follows that $[x]_\Phi = [y]_\Phi$. Therefore every pair of distinct equivalence classes has empty intersection and we deduce that the family of all equivalence classes is a partition, $\mathbf{P}(\Phi)$, of the set A.

There are some very interesting examples of equivalence relations. First let M be the set of all sequences $\mathbf{s} = (x_n)_{n\in\mathbb{N}}$ of rational numbers. Consider the relation Φ on M defined by the rule: $(\mathbf{s},\mathbf{r}) \in \Phi$ if and only if

$$\lim_{n\to\infty}(x_n - y_n) = 0.$$

Here $\mathbf{r} = (y_n)_{n\in\mathbb{N}}$. It it easy to see that Φ is an equivalence relation.

For another example, let $M = [0,1]$. Define a relation P on M by $(x,y) \in P$ if and only if $x - y$ is a rational number. It it easy to see that P is an equivalence relation.

There is one more important **example**.

Let m be a fixed natural number. Two integers are called congruent modulo m if $a - b$ is divisible by m, which we denote by $a \equiv b\,(\mathbf{mod}\ m)$. This congruence relation is easily shown to be an equivalence relation, which we shall consider in detail later.

We often denote equivalence relations using symbols such as $\cong$, $\equiv$, $\approx$, $\sim$, and others.

Exercise Set 1.5

In each of the following questions explain your reasoning, either by giving a proof of your assertion or a counterexample.

1.5.1. On the set $G = \mathbb{Z}\times\{-1,1\}$ we define an operation $*$ by the rule $(m,a) * (n,b) = (m + an, ab)$. Is this operation associative? Commutative? Is there an identity element? Which elements have inverses?

1.5.2. On a set of four elements define a commutative, associative binary operation having an identity element.

1.5.3. On the set $\mathbb{Z}$ define an operation $\perp$ by the rule $a\perp b = a^2 + b^2$, for $a,b \in \mathbb{Z}$. Is this operation associative? Commutative? Is there an identity element?

1.5.4. On the set $\mathbb{R}$ define an operation $\bullet$ by the rule $a \bullet b = a + b + ab$. Prove that

(i) $a \bullet (b \bullet c) = (a \bullet b) \bullet c$ for all $a,b,c \in \mathbb{R}$.

(ii) $a \bullet b = b \bullet a$ for all $a,b \in \mathbb{R}$.

(iii) if $a \neq -1$, then $a \bullet b = a \bullet c$ if and only if $b = c$.

Is there an identity element for this operation? Which elements have inverses?

1.5.5. On the set $\mathbb{R} \times \mathbb{R}$ define an operation $\bullet$ by the rule $(a,b) \bullet (c,d) = (ac-bd, bc+ad)$. Is this operation associative? Commutative? Is there an identity element?

1.5.6. Let $M = \{e,a,b,c\}$. Define a binary algebraic operation on M which is commutative, associative, and for which an identity element exists, but not every element has an inverse.

1.5.7. Let $M = \{e,a,b,c\}$. Define on M a binary algebraic operation which is commutative, associative, and for which there is an identity element, and every element has an inverse.

1.5.8. For $a,b \in \mathbb{R}$ define $a \simeq b$ to mean that $ab = 0$. Prove or disprove each of the following:

(a) The relation $\simeq$ is reflexive.
(b) The relation $\simeq$ is symmetric.
(c) The relation $\simeq$ is transitive.

1.5.9. For $a,b \in \mathbb{R}$ define $a \simeq b$ to mean that $ab \neq 0$. Prove or disprove each of the following:

(a) The relation $\simeq$ is reflexive.
(b) The relation $\simeq$ is symmetric.
(c) The relation $\simeq$ is transitive.

1.5.10. Fractions are numbers of the form $\frac{a}{b}$ where a and b are whole numbers and $b \neq 0$. Fraction equality is defined by $\frac{a}{b} = \frac{c}{d}$ if and only if $ad = bc$. Determine whether fraction equality is an equivalence relation.

1.5.11. Let $a,b \in \mathbb{N}$. Show that the operation $a * b = a^b + b^a$ is not associative on the set of natural numbers. Is it a commutative operation?

1.5.12. Let M be the set of sequences $\mathbf{s} = (x_n)_{n\in\mathbb{N}}$ of rational numbers and let also $\mathbf{r} = (y_n) \in M$. Prove that the relation Φ defined by $\mathbf{s}\,\Phi\,\mathbf{r}$ if and only if $\lim_{n\to\infty}(x_n - y_n) = 0$ is an equivalence relation on M. Write the first few terms of a sequence defining π.

1.5.13. Let m be a fixed natural number. If $a,b \in \mathbb{Z}$ then write $a \equiv b \pmod{m}$ if and only if m divides $b-a$. Prove that $\equiv$ is an equivalence relation on the set $\mathbb{Z}$. Write the equivalence class of $0 \pmod 7$ and the equivalence class of $0 \pmod 5$. Are these the same?

1.5.14. Prove that the relation "has the same image" on the elements of the domain of a function is an equivalence relation.

1.5.15. Define binary operations ▼, ▲ and ■ on $\mathbb{Q}$ by the rules:
$a\blacktriangledown b = a - b + ab, a\blacktriangle b = \frac{1}{2}(a+b+ab)$, $a\blacksquare b = \frac{1}{3}(a+b)$.
Of these operations which are associative? Which are commutative? Which have an identity element?

1.5.16. Define a binary operation ▼ on $\mathbb{R}$ by the rule:
$a\blacktriangledown b = pa + qb + r$. For which fixed p, q, r, is this operation associative? For which values of p, q, r is the operation commutative? For which values of p, q, r is there an identity element?

1.5.17. Let $\mathbb{Q}^*$ be the set of all nonzero rational numbers. Which of the following properties hold for the operation of division:

(1) $a \div b = b \div a$;
(2) $(a \div b) \div c = a \div (b \div c)$;
(3) $((a \div b) \div c) \div d = a \div (b \div (c \div d))$;
(4) if $a \div b = a \div c$, then $b = c$;
(5) if $b \div a = c \div a$, then $b = c$.

1.5.18. For $a, b \in \mathbb{R}$ define $a \simeq b$ to mean that $|a - b| < 7$. Prove or disprove each of the following:

(a) The relation $\simeq$ is reflexive.
(b) The relation $\simeq$ is symmetric.
(c) The relation $\simeq$ is transitive.

1.5.19. For points $(a, b), (c, d) \in \mathbb{R}^2$ define $(a, b) \simeq (c, d)$ to mean that $a^2 + b^2 = c^2 + d^2$.

(a) Prove that $\simeq$ is an equivalence relation on $\mathbb{R}^2$.
(b) List all elements in the set $\{(x, y) \in \mathbb{R}^2 \mid (x, y) \simeq (0, 0)\}$.
(c) List five distinct elements in the set $\{(x, y) \in \mathbb{R}^2 \mid (x, y) \simeq (1, 0)\}$.

1.5.20. Two $n \times n$ matrices A and B are said to be similar if there exists an invertible $n \times n$ matrix P such that $P^{-1}AP = B$. Show that similarity is an equivalence relation on $\mathbf{M}_n(\mathbb{R})$.